

## When to publish pseudo-science

*The opinion that this journal should not have published a famous article on solute-free solutions overlooks the multivalent function of a general journal of science.*

NATURALLY, but mistakenly, a handful of readers (and some others) are perplexed that *Nature* should have published the article by Dr Jacques Benveniste and his associates (333, 816; 1988) reporting the unbelievable circumstance that solutions of a biological reagent retain their activity even when diluted beyond the point at which molecules of the active ingredient can persist. One view, ably put by Dr Henry Metzger on page 375, is that the publication of irreproducible data, as distinct from speculative opinion, requires that more exacting editorial standards should apply. Otherwise, he says, the process of science is confused, even lent a "circus atmosphere".

Metzger goes on to echo a not disinterested toffee-nosed opinion recruited last week by the *New York Times* that journals such as this should not lend their reputation to spurious science by publishing it; Metzger uses the word *imprimatur*.

Sadly, real science is less perfect and the truth is more complicated. General journals such as this are afflicted by a torrent of heterodox would-be literature offered for publication. Proofs of the error of special relativity are the stock of this trade, but novel theories of mind are catching up. Deciding what to do is usually simple, which is not to imply that dealing with these unpublished manuscripts is straightforward; the authors are usually zealous correspondents — which, given the enormity of their claims, is natural enough. But it is rare that some such claim should come from a government-supported laboratory, that its principal author should urge publication in the face of common sense — and should complain that failure to publish will be tantamount to the suppression of the truth. The antecedents of the Benveniste publication are as exceptional as its sequel.

Moreover, there is no absolute rule that journals such as this, which are proud to publish a great deal of original science, should not on occasion give an airing to material that is different in kind. Indeed, there are occasions when publication of spurious science may be a public service. Some readers may recall the case of scotophobin, a protein supposed to reside in the brains of trained rats which, when injected into the brains of naive rats, would transfer the first rat's learned capacity to run a maze, for example. *Nature* published a version of such a manuscript after several preliminary accounts had appeared elsewhere, but accompanied it with a devastating critique from Mr Walter W. Stewart (see *Nature* 238, 198–210 1972). Nothing much has been heard of scotophobin since. Is not a little of the "circus atmosphere" inescapable on these occasions?

Not that belief in the magical properties of attenuated solutions will be as quickly exorcised. Since the emergence of homoeopathic medicine in the early nineteenth century, before the atomic theory had been properly established, the theory of biological activity at extreme dilution has been a theory in search of verification. It would be naive to expect that the hunt for verification will now be abandoned simply because *Nature's* opinion of Benveniste's experiments is unsatisfactory. When a quarter of French physicians prescribe homoeopathic medicines, there is plainly too much at stake for the issue to be dropped. This journal will keep an eye on this interesting field, with its implications that the law of mass action must be repealed. That is consonant with the function of a general journal of science.

Strictly academic journals have simple rules to follow when deciding what to publish. General journals such as this, whose function is not only to publish excellent original science but to keep a general readership well-informed, necessarily have other fish to fry as well. Merely reporting events will not always suffice. But the two functions are complementary, for who can pretend, alas, that excellent science is the whole of science?

## When less means less

*Britain plans to spend less and less on research, but has not explained why.*

UNDERSTANDING British policy on research has become a task comparable with that of guessing what may lie behind the latest moves by one's opponent in a game of chess. The game the British government is playing on the financing of research may have the admirable quality of consistency, but its objectives are entirely obscure. For chess-players, that is a splendid posture; the interesting question is whether it can be appropriate in the conduct of public policy. Two documents appearing last week illustrate why it is not.

First, there is the now-annual review of the Advisory Board for the Research Councils (ABRC), this year as always a repetition of Oliver Twist's familiar theme of "More, please!" (see *Nature* 334, 279; 1988). As the years go by, the tone of this demeaning request is increasingly despondent.

Again, much of the board's need of extra cash stems from its wish, consistent with government ambitions, to assist with the "restructuring of the science base", which is another name for the further amalgamation of institutes and laboratories. The request is repeated partly because last year's extra cash was swallowed up by nationally negotiated increases of university pay. Yet ABRC also notes the continuing decline of British contributions to the international literature of science, again zealously recorded by John Irvine and Ben Martin from the University of Sussex. Actual spending through the research council system, £699 million this year, is planned to increase by 0.4 per cent between now and next year (when inflation is likely to exceed 5 per cent a year) and, thereafter, to fall in cash terms.

Why should Britain be cutting back on research when governments elsewhere are doing the opposite? Sadly, there seems to be no answer. But the annual review by the Cabinet Office of public support for research and development (see page 374) faithfully catalogues the downward trend — a decrease of 3 per cent between 1985 and 1986 will be followed by a further decrease of 7 per cent by the end of the decade. The only good thing to be said for this pattern is that defence research will bear the brunt of the decline.

Oddly, no explanation is given. The most plausible theory is that the British government has decided that its own economies on research will force industry to spend instead. In the past year, there has indeed been an increase of the total volume of research spending of £250 million, but more than 70 per cent of the extra was spent by the British subsidiaries of overseas corporations. That seems to be a high-risk strategy. □



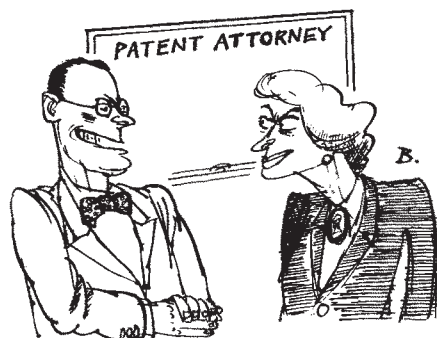
# Lead in superconductivity patent race to go to Japan?

□ Nobel prizewinners running second

□ Patents filed on three continents

Tokyo

THE Japanese patent office revealed on 20 July that Tokyo University has beaten International Business Machines (IBM) Co. off the mark in filing for a patent that describes the composition and structure of lanthanum-barium-copper oxide, the high-



At near-normal temperatures, these new superconductors can keep a supply of money going indefinitely...

temperature superconductor discovered by Nobel prizewinners Georg Bednorz and Alex Müller at IBM Zurich in 1986.

The patent application is the first of its kind to be made public and it draws attention to differences between Japanese, European and US patent laws.

A team at Tokyo University led by Professor Shoji Tanaka filed the application for the patent in Japan on 17 January 1987 on the basis of its confirmation of the IBM discovery and, more important, its determination that the structure of lanthanum-barium-copper oxide is the same as that of crystals of  $K_xNiF_4$ . According to Tanaka, IBM Zurich filed a "very similar" application in Europe six days later.

Michael Starks, an IBM spokesman on patents, confirmed that such a European patent had been filed. And he says that IBM has also applied for high-temperature superconductor patents in the United States and Japan. The Tokyo University patent was made public under Japanese law which requires that all patents are published 18 months after application to allow objections to be raised before the patent is awarded as in Europe.

As both Japan and Europe determine patent priority on the basis of who is 'first to file', the Tokyo University patent has priority, at least in Japan (and also in Europe if a patent application has also

been filed there on the basis of the Japanese application).

This contrasts with the situation in the United States where the 'first to invent' has priority. But the US rule applies only to research carried out by researchers based in the United States or temporarily assigned overseas for research on behalf of the United States. And IBM cannot invoke this rule even for a US patent because its research was carried out in Zurich by non-US citizens.

IBM could file an objection in Japan stating that the Tokyo University patent application is 'obvious' because the superconductor was described by Bednorz and Müller in a scientific publication in September 1986. But, if successful, this course of action would bar both Tokyo University and IBM from getting a Japanese patent, according to Yasunori Ohtsuka, a patent attorney in Tokyo. And such action would also run counter to IBM's claims for the European patent.

According to Tanaka, both his patent and IBM's European patent application were made on the grounds that the crystal structure of the superconductor was not described in the September 1986 paper.

Under European patent law, applications for patents cannot be made for discoveries that have already been published. This contrasts with the United States, where patent applications can be filed within one year after publication, and Japan where six months is allowed.

But the real battle over superconductor patents has yet to begin. Professor C. W. Chu filed an application for a US patent for yttrium-barium-copper oxide within days of the Tokyo University application. The yttrium-based copper oxides superconduct at higher temperatures than those made from lanthanum and barium and applications of the yttrium-based material are much more advanced.

Unlike Japanese patents, US patent applications remain confidential until granted. And it is only after they are awarded that they can be contested. Chu's patent is expected to be awarded very soon. And hundreds of Japanese patents for applications of the new superconductors are also expected to be revealed by the Japanese patent office over the next few months. Clearly patent lawyers around the world are going to have their hands full in determining priority when this barrage of patent applications is made public.

David Swinbanks

## François Kourilsky to head CNRS

Paris

THE French Research Ministry has named François Kourilsky, a 53-year-old biologist, as director of the Centre National de la Recherche Scientifique (CNRS). Kourilsky fills the vacancy left after Serge Feneuille unexpectedly resigned last month (see *Nature* 334, 5; 1988).

If it is unusual for a biologist to head France's largest public-sector research organization (a long line of physicists have previously had the job), Kourilsky's pedigree makes him a natural choice to accompany Hubert Curien's second tour as Research Minister in the new Mitterrand government. From 1983 to 1987, Kourilsky was vice-president of the Conseil Scientifique de la Recherche et de la Technologie (CSRT), the government-appointed committee of academics and industrialists that advises the ministry on research policy. Curien, as research minister, is its president.

The CSRT was set up under the previous socialist government as a symbol of *glasnost* and took an active part in policy-making, recommending annual job creation for researchers and a steady budget progression. But, when the socialists lost power in 1986, the CSRT became something of a white elephant and was largely circumvented (see *Nature* 328, 567; 1987). Although the government has a statutory obligation to submit its research budget proposals to the CSRT for comment, the 1988 science budget was shown to the committee only a few hours before it was announced to the press. Kourilsky's appointment is therefore yet another sign of President Mitterrand's wish to pick up where he left off in 1986, despite promises to let the political centre have more say.

Kourilsky is well known in the field of immunology and, last year, set up an AIDS research unit with Jean-Claude Chermann at Marseilles, with funding from the Institut National de la Santé et de la Recherche Médicale (INSERM). After a stormy time under the previous government, CNRS researchers will welcome Kourilsky's appointment. But, according to a Research Ministry communiqué, he has been asked to "adapt and modernize the functioning of this large organization, making it more supple, while taking account of the predominant role of the personnel and their motivations". This task has fallen to successive directors, none of whom has been able to make significant changes.

Peter Coles

### Erratum

In last week's account (pages 287-290) of our investigation of the claims of Dr Jacques Benveniste and his associates, Fig. 3 was cited in place of Figs 4 and 5, and should have been cited at the bottom of the first column on page 289. □



# US turns blind eye to untested AIDS treatment drug imports

## Washington & Tokyo

US Food and Drug Administration commissioner Frank Young announced last week that the government would no longer try to stop the import of unapproved drugs for the personal treatment of AIDS.

The decision will affect most strongly the import of dextran sulphate, a drug manufactured in Japan for the treatment of high blood-lipid levels and shown to hinder HIV virus binding in cell culture (see *Nature* 334, 3; 1988). The drug has become a popular self-treatment for AIDS patients in the United States. Imports by mail will now be permitted provided the amount is no more than a three-months' supply.

Young's announcement, made to a hostile audience at the National Lesbian and Gay Health Conference, represents a capitulation to the overwhelming pressure from people with AIDS to be allowed to try to treat themselves with drugs that are still untested. Only one drug, AZT, has received FDA approval for the treatment of AIDS and it has such powerful side-effects that many AIDS patients have abandoned therapy with it.

Until last week's announcement, it appeared that the government was preparing to take a tougher line to restrict the availability of untested remedies. Earlier in the year, dextran sulphate, which is manufactured only in Japan, had been freely available from "buyers health clubs" in San Francisco, Los Angeles and New York and could be ordered by mail, even though it has not been tested for toxicity in the United States. But in April, in response to complaints from the United States, Japan began to make it harder for foreigners to buy the drug and export it.

Protests followed in the United States, including sit-ins by AIDS activists' groups in the Japanese consulate in San Francisco and the Manhattan office of Kowa, one of the biggest manufacturers of dextran sulphate.

The Japanese Ministry of Health and Welfare will now allow three pharmacies — two close to Tokyo's international airport at Narita, and one, the American Pharmacy, close to the Ginza shopping district — to sell the drug to foreigners for export. A spokeswoman at the American pharmacy said that they would sell a maximum of 10,000 300-mg tablets to any one individual. The purchaser must show a passport and sign a form recognizing the drug has no proven effect against AIDS.

Free availability of dextran sulphate may make it harder to test new therapies against AIDS. Many doctors running clinical trials report that their patients

were taking unknown quantities of other untested drugs. In tests of AZT, some patients took their medicine to independent laboratories for analysis; those who found they had been given a placebo dropped out of the trials, making assessment of the treatment more difficult. Interpretation of phase I toxicity tests of dextran sulphate were also complicated when patients boosted their hospital doses with dextran sulphate bought from health clubs.

A spokesman at the San Francisco

General Hospital, where dextran sulphate has been under test, stressed that patients involved in clinical trials "had to be a little altruistically motivated", "trust their doctors" and recognize "that they would do themselves and other people with AIDS a disservice by taking untested drugs". But people-with-AIDS groups say that trust has been hard to give when it appears that the government is not testing and licensing new drugs quickly enough and is insensitive to the need of those dying from AIDS to try anything that might help.

The new official policy on untested therapies may help to restore faith in the FDA's good intentions.

Alun Anderson & David Swinbanks

# House of Representatives due to decide on animal patents

## Washington

THE US House of Representatives Committee on the Judiciary will this week consider two drastically different ways to resolve the complex ethical, procedural and economic issues surrounding the patenting of animals. The committee will decide whether Congress should establish a two-year moratorium on the granting of patents covering animals, or allow the patenting of animals to continue, but exempt researchers and farmers from paying royalty fees after breeding the animals.

Congress became obliged to set policy over the patenting of animals after the US Patent and Trademark Office decided in April of last year to accept applications for animal patents. The patent office ruled that it could no longer stand in the way of patenting animals simply because they were higher life forms (*Nature* 326, 729; 1987). During the public furor that then erupted, the patent office made clear that its hands were tied by the judicial interpretation of current patent laws, and that it was up to Congress to weigh the implications and set down new laws if the policy was unacceptable.

For the past year, animal rights activists, religious leaders, farming organizations and anti-biotechnology agitator Jeremy Rifkin have formed unlikely alliances to lobby Congress to assess the patent office's new policy. Senator Mark Hatfield (Republican, Oregon) and Representative Charles Rose (Democrat, North Carolina) both entered legislation last year to suspend the granting of patents on animals until the surrounding issues could be evaluated.

But while Congress has been in a flurry of activity to come to grips with the technology to create "new" animals, the patent office has continued to process the roughly 20 applications it has on hand. It

granted a patent to Harvard University in April when all of the legal hurdles to patentability were jumped by "myc-y mouse" — a mouse made susceptible to the development of cancer by the insertion of the *myc* oncogene into its genome (*Nature* 332, 668; 1988).

Congress this week is reconsidering Representative Rose's bill calling for a two-year moratorium on the patenting of animals, along with a more moderate bill entered by Representative Robert Kastenmeier (Democrat, Wisconsin) which would exempt researchers and farmers from the royalty payments required under normal patent law for the use of patented animals. No action has yet been taken on a stricter bill proposing an indefinite moratorium on animal patenting which has been entered again this year in the Senate by Senator Hatfield.

Representative Kastenmeier has also sponsored a bill to establish a transgenic animal advisory committee in the Department of Agriculture to oversee the environmental release of genetically engineered animals eligible for patenting.

Critics of the patenting of animals state that it equates "cows with toasters", and that it will lead to higher prices for farm and laboratory animals and reduce the genetic diversity of animals important to US agriculture. Proponents claim that it encourages innovation, and contend that the patenting of plant varieties — permitted since 1970 — has not been the cause of a recent escalation in seed prices.

But some are not prepared to wait for Congress to resolve these issues: the Animal Legal Defense Fund, an animal rights organization, last week filed suit against the US patent office accusing it of violating the Patent Act, and the laws governing the establishment of new policies by Federal agencies, by granting animal patents.

Carol Ezzell



# US faces danger in living off the fat of the land

## Washington

AN encyclopaedic, 712-page report\* on the links between diet and disease received a mixed reception last week when it was released by the US Surgeon General, C. Everett Koop. Critics argued that the report, which took four years to prepare, breaks no new ground and fails to spell out specific nutritional guidelines. But the report's authors defended it as a comprehensive account of current knowledge from which policy recommendations and public programmes will later flow.

Koop described the report as a "landmark in Public Health Service efforts to improve the health of the American people" and compared it with the report on smoking, by Surgeon General Luther Terry in 1964, that triggered public recognition of the dangers of cigarette smoking.

In *Nutrition and Health*, the major villain is fat, particularly saturated fat (the kind, like butter, that remains solid at room temperature). Americans eat too much fat, according to Koop, and as a result suffer from increased rates of coronary heart disease, some types of cancer, strokes, diabetes and atherosclerosis. In Mediterranean countries, Japan and China, where much less fat is consumed, heart disease rates are much lower than those in the United States.

Coronary heart disease is now the United States' biggest killer. Of the 5,800 Americans who die each day, 35.7 per cent will succumb to heart disease, 22.4 per cent to cancer, 7.0 per cent to strokes, 1.8 per cent to diabetes and 1.1 per cent to atherosclerosis.

With twice as many calories per gram as carbohydrates and proteins, fat is also very fattening. Excessive consumption has led to an "extraordinarily high prevalence of obesity". One in four American adults are overweight and one in ten are severely overweight. Obesity brings increased risk for diabetes, hypertension, coronary artery disease and strokes, gall-bladder disease and some types of cancer.

The surgeon general's advice is to switch from red meats and dairy products to lower-fat foods: vegetables, fruits, whole grains, fish and poultry. And he recommends an increase in consumption of complex carbohydrates and fibre grains and cereals, vegetables and fruits. Epidemiological and animal studies suggest they help to reduce the risks of cancer and diverticulosis, and may reduce blood cholesterol levels.

Attempts by Americans to improve their health by eating \$1,500 million per year of extra vitamins, minerals and other dietary supplements receives short shrift in the report. They are judged generally

unnecessary, and excess consumption of them to be hazardous. Judgement is suspended on Omega-3 fatty acids from fish oils, an increasingly popular health-store product. Evidence that it reduces blood cholesterol levels and heart disease risk is considered too preliminary to recommend changes in intake.

Excess consumption of alcohol and salt are linked to increased mortality rates. Alcohol is seen as a factor in automobile and other accidents, and in suicides. These are the fourth (4.4 per cent) and

eighth (1.4 per cent) most frequent causes of death in the United States. The surgeon general recommends moderation — "no more than two drinks a day, if at all". Excess salt consumption is linked to high blood pressure and strokes. Average sodium intake for adults in the United States is 4–6 grams, four times the level judged safe and adequate.

To get across the report's message, Koop has called for a national campaign to change the American diet. But he stresses that the report is largely aimed at policy-makers. They will have to decide how it may be turned into action. **Alun Anderson**

\*The Surgeon General's Report on Nutrition and Health (US Department of Health and Human Services, Public Health Service.)

# Vital boost for rocket fuel

## Las Vegas & Boston

LAST week, the Pacific Engineering and Production Company (PEPCON) unveiled plans for a new rocket fuel plant to replace the one destroyed in a gigantic explosion in Nevada nearly four months ago. The plant will manufacture ammonium perchlorate, a crystallized, oxidizing agent that constitutes more than two-thirds of the make-up of solid rocket fuel. The announcement has been eagerly awaited by the military, the National Aeronautics and Space Administration (NASA), Congress and the commercial space industry, because a new facility is vital if a shortage of rocket fuel is to be avoided next year.

The cause of the accident at the PEPCON plant in Henderson, Nevada, is still being investigated, but preliminary reports have attributed it to a gas leak under the plant. The explosion demolished the plant and halved the United States' production capacity of ammonium perchlorate. Two people were killed, hundreds were injured and millions of dollars worth of damage was done to the adjacent town. Local legislators are now trying to push through tough new laws governing the manufacture of hazardous chemicals.

At the national level, the PEPCON blast raised fears that an impending rocket fuel shortfall could cripple the US space launching capacity. The result has been a flurry of activity among US military, NASA, congressional subcommittees and the private companies involved.

To speed plans to increase production capacity, the Pentagon and NASA have pledged to buy a minimum of 100 million pounds (weight) of ammonium perchlorate over the next five years from both US producers. According to government estimates, 62 million pounds of ammonium perchlorate were being produced annually before the explosion. Maximum capacity is now under 40 million pounds and demand is expected to rise from approximately 58 million to 70 million pounds per year over the next few years. Although 24 million pounds is believed to be stockpiled, a serious shortfall may develop late next year.

Now that the location of a new facility has been established, PEPCON claims it can be back in production by February 1989, but others are doubtful. As NASA official Russell Bardos told reporters last month, "it will be 1991 before we can breathe easy again". **Seth Shulman**



*The centre-piece of the Soviet Union's 'Atoms for Peace' exhibition in the autumn, Pravda reports, will be a statue by the sculptor Nikolai Selivanov of Igor Kurchatov, pioneer of Soviet nuclear energy, Sergei Korolev, the rocket designer, and Mstyslav Keldysh "one of the initiators of the opening up of space". The link between nuclear power and space would be questioned by Academician Roald Sagdeev, who wants a ban on nuclear reactors in Earth orbit.*

# HOTOL seeks support abroad as Britain opts out of space

## London

THE British space community was dealt a severe blow last week when the government refused funds for the development of the space plane Hotol, the revolutionary horizontal take-off and landing space vehicle developed by British Aerospace (BAe) and Rolls Royce. They had requested £12 million over the next three years to take the project beyond the proof-of-viability phase which was successfully completed last year. That phase was funded jointly by the government and industry, each providing £1.5 million. The designer of Hotol's unique engine, Alan Bond, says

dominate the world, says Bond. Britain would be in a position to challenge that domination only if development of Hotol went ahead bringing corresponding advances in space engineering, says Bond.

BAe has now intensified attempts to find foreign partners, but progress is hindered by classification by the Ministry of Defence of Hotol's unique propulsion system. Until now the government has been very reluctant to declassify Hotol. Now that the government has said it will support efforts to find collaborators, it is considering declassification. Without that it will be very difficult, though not impos-



The British space plane that could make cheap space trips a reality.

there is no doubt now that the space plane is technically feasible. "It could be on the runway in 12 years" he says. And it would reduce by about one half the cost of launching satellites. At present, it costs about £13 million to put a one-tonne payload into geostationary Earth orbit; Hotol could do it for between £6 and £8 million.

The future of the spaceplane now depends on finding sponsors for its development. It was never intended that the programme should be purely national; BAe has for several years been seeking foreign partners. But the government's withdrawal seems to be a vote of no confidence in the project. "This has very seriously jeopardized the future of Hotol", says Bond. "Getting 100 per cent support from abroad will be very difficult. Countries which might be expected to collaborate are quite capable of going ahead with similar projects themselves". Britain is leading the world now, he says, but the rest of Europe is only two or three years behind. And Japan and the United States, now about 12 years behind, are making rapid advances.

If Britain misses out on Hotol, then it will quickly fall behind the international space community, says Bond. Hotol fills, at an economic price, a gap between conventional rockets and the aerospace plane of the late twenty-first century. But when increased traffic makes development costs less significant, then the expensive US propulsion system, SCRAMJET, will

sible, says Bond, to elicit the necessary foreign support.

That support is vital both financially and in terms of the expertise it would bring to the project. The skills to develop the space plane no longer exist in Britain, says Bond. "At most we could expect to do 35 per cent of the work. Successive governments have run us into a state which is shameful". Bond believes that the Hotol programme brought together a team of British specialists that is world-class. That expertise is in danger of slipping away.

Though development costs may be high — between £5,000 and £7,000 million — profits are not as distant as people think, says Bond. Within about 20 years, Hotol would be making a profit. And although the space plane has been designed for launching communications satellites, it can be modified to take any payload of up to eight tonnes and with further development could put astronauts into space, and even act as a commercial passenger-carrying vehicle.

The government's withdrawal from Hotol depressed but did not surprise the British space community, which is already demoralized by the government's consistent refusal to increase the space budget from the current level of £130 million. And last year the head of the British National Space Centre resigned after his attempts to formulate a national policy were continually frustrated by refusal of public funding.

Christine McGourty

# Leaks prompt delay in shuttle launch

## Washington

LEAKS continue to plague preparations for the forthcoming launch of the space shuttle Discovery. Last weekend the National Aeronautics and Space Administration (NASA) completed a practice countdown that was to include filling the shuttle's fuel tanks with liquid hydrogen and liquid oxygen fuel. But during the fuelling procedure a leak developed in a hydrogen line leading from the fuel tanks on the launch pad to the shuttle fuel tank, and NASA abandoned plans for a complete fuelling. Earlier in the test a problem cropped up with a liquid oxygen recirculation pump, also part of the shuttle's ground support equipment.

NASA will decide later this week when to proceed with a 20-second test firing of the shuttle's main engine. The results of that test will help determine whether engineers will attempt to repair at the launch pad a leak discovered last month in Discovery's on-orbit positioning system, or whether it will be necessary to roll the shuttle back to its hangar. If a rollback is necessary, the launch would be delayed by approximately two months. In any case, a launch before October is now problematic. Joseph Palca

# Concession on education bill

## London

THE UK government last week made a surprise last-minute concession to academics over the controversial clause in the education reform bill on the 'terms and conditions' which the new University Funding Council (UFC) can lay down on universities (see *Nature* 334, 284; 1988). Fears that the clause would allow the introduction of restrictive contract funding led to an amendment in the House of Lords which made the powers of the council less specific. But that was subsequently overturned in the House of Commons. To appease dissatisfied peers, the government quickly introduced a new clause which imposes a statutory duty on the UFC to consult appropriate bodies before imposing any 'terms and conditions'. Those bodies would be representative organizations where the whole sector is concerned, such as the Committee of Vice Chancellors and Principals (CVCP), or the governing body where an individual institution is concerned. Peers consented and the bill duly became law on 29 July, the end of the parliamentary session. Sir Mark Richmond, chairman of the CVCP, says the late clause is a "substantial extra concession". And the final form of the bill, he says, is "spectacularly improved" in respect of the universities on the original form of the bill.

Christine McGourty



## Benveniste controversy rages on in the French press

### Paris

THE publication in *Nature* (333, 816–818; 1988) of Jacques Benveniste's claims for the continued biological activity of solutions no longer containing molecules of the active constituent provoked a flurry of excitement in the French press. The absence of effect subsequently found by *Nature*'s own team of investigators (see *Nature* 334, 287–291; 1988) has, with almost poetic irony, created at least as much of a press reaction as the original claim. Meanwhile, if Benveniste's own institute, INSERM (Institut National de la Santé et de la Recherche Médicale) is embarrassed by the controversy, Laboratoires Boiron, the manufacturers of homeopathic products which co-sponsored Benveniste's work, have been busy on the stock market.

Benveniste is no stranger to the French press. While *Nature*'s editors were still considering whether to publish the research, the French newspaper *Le Monde*, in its issue dated 29–30 May 1988, interviewed Benveniste after he had presented his findings to a national homeopathy conference in Strasbourg. Admitting the results could not be explained by current theory, Benveniste threw down the gauntlet, telling *Le Monde* that "only the publication of our work in an undisputed international journal will enable us to advance".

When the paper was published, it caused reactions in the French media ranging from 'incredible-but-true' to fantastic speculations. While *Le Figaro* has maintained a reserve throughout, balancing Benveniste's comments with the respectful disbelief of Nobel prizewinner Jean-Marie Lehn, *Le Monde* and *Libération* have joined the mud-slinging. Between the publication of Benveniste's paper and last week, *Le Monde*, with an opening sentence "magic to the rescue of reason", demoted *Nature* with the phrase "until now one of the most prestigious [scientific journals] in the world".

*Libération*, which pre-empted *Nature*'s publication of its report by giving Benveniste's own account of the inquiry, also felt, in a three-page follow-up on July 29, that *Nature* "has perhaps more to lose than Benveniste", saying *Nature*'s editors were "the most two-faced of the lot".

If the choice of James Randi and Walter Stewart as arbiters has scandalized Benveniste, the attitude of the INSERM directorate has also wounded him. In an initial communiqué, dated 29 June, INSERM refused to express an opinion on the results. "Every real discovery inevitably incites temporary incredulity, using its customary methods of evaluation,

where scientific controversy has its place, to sort out what, in the final account, turns out to be no more than illusion, from that which constitutes a real advance in knowledge."

When *Nature* subsequently published its rebuttal, INSERM issued another statement, again saying it would not enter the debate. "The institute repeats that all of its laboratories are themselves responsible for their choice of research themes and are, every four years, subjected to an in-depth scientific evaluation of their activities. Unit 200 of INSERM, directed by Dr Benveniste, will be subjected to its next regular examination in Spring 1989. It is within this framework that it will fall to Dr Benveniste's peers to make their judgement on the entirety of the work of the group, at a time when we hope that the understandable passions released or maintained by the to-and-fro of the two successive publications in *Nature* will have been calmed, leaving room for the indispensable serenity of scientific judgements in the long run."

In the 29 July issue of *Libération*, Ben-

veniste complained that INSERM had, effectively, thrown him to the wolves. "*Nature* sends a magician to check my research and INSERM doesn't even protest. It's the limit!"

While the scientific controversy continues, the industrial co-sponsors of Benveniste's research, Laboratoires Boiron, have been making sure they benefit from any renewed interest in homeopathy. Boiron, already a 51 per cent shareholder in another major manufacturer of homeopathic medicines, Laboratoires Homéopathiques de France (LHF), last week decided to buy up the remaining shares. In *Le Monde* of 29 July, a Boiron spokesman is reported as saying that "*Nature* is trying to cover itself...and must be under great pressure to attenuate the impact of the recent publication. One must be aware that the scientific community is faced with a reality that has 'put the wind up it' and that it will react in a variety of ways".

The last word (so far) goes to the monthly popular science magazine, *Science et Vie*, which claims to have proved that Benveniste's results demonstrate only the discoloration (*achromasie*) of the toluidine stain, not degranulation of basophils, also wondering: if water has a 'memory', how can it ever be 'pure'?

Peter Coles

## Restructuring of UK research could be misguided

### London

THE academic research system in Britain is being restructured in order to get the best value for money at a time when government policy is constricting public funding. But the current restructuring policy which grades individual department and not whole institutions may be inappropriate, according to a study published last week by John Irvine of the Science Policy Research Unit at the University of Sussex. In his study of the performance of 13 technology-oriented universities, he finds that research excellence in a broad range of fields is concentrated in a small number of institutions.

His findings seem to support the view of the government advisers on science policy, the Advisory Board for the Research Councils (ABRC) that research resources should be concentrated in centres of excellence. The ABRC proposed last year that higher education institutions should be divided into three types: research-based, mixed research and teaching, and teaching based. But that proposal was severely criticized by the academic community.

The University Grants Committee, adviser to the government on the distribution of grants, then proceeded with its policy of restructuring universities on a subject-by-subject basis, with a view to

making recommendations for support to individual departments regardless of the overall performance of host institutions. A review of the Earth sciences was completed this year and that too was criticized by academics. Reviews of chemistry and physics are now under way and preliminary reports are due soon.

Irvine based his study on comparisons of science citations and the influence of the journals in which the citations appear. Of the institutions studied, Irvine says preliminary results show strong indications of concentration of research excellence, with three universities being in the top five for chemistry, engineering, physics, biological sciences, mathematics and Earth and space sciences.

The higher classified universities receive a proportionately greater amount of research grants and a higher level of contracts and grants from companies, says Irvine. And he finds a strong linkage between output of high quality basic scientific and engineering research and its perceived utility to industry.

Irvine acknowledges that bibliometric studies of research performance are subject to limitations, and says further research is necessary before reliable conclusions on which to base policy decisions can be drawn.

Christine McGourty

# Soviet worries about nuclear safety after Chernobyl

## London

PLANS for the Chernobyl nuclear power station have been changed, Soviet television news announced last week. The original plans for the rehabilitation of the power station after the disaster included not only the recommissioning of Blocks 1 and 3, but also the completion of blocks 5 and 6, which were under construction at the time of the accident. Now, however, it has been decided to dismantle blocks 5 and 6 instead.

This is not entirely surprising. In March 1986, a month before the accident, the weekly *Literaturna Ukrayina* carried an alarming account of skimped work and corner cutting in the construction of block 5. Last week's decision seems to have been based, however, not on the possible defects on blocks 5 and 6 but on the whole issue of the safety of the RBMK reactor. After the disaster, the Soviet nuclear energy authorities committed themselves to phase out the RBMK programme, and to introduce additional safety systems on those reactors already in service. During the past two years, however, public anxiety about the RBMK has grown. Moreover, it has now emerged that, from the very beginning, some Soviet scientists had considerable doubts about the safety of the RBMK, but were overruled by the "scientific councils" in charge of decision-making. "Dissenting opinions required", should be written over the doors of all nuclear power institutes, commented *Literaturnaya Gazeta* two weeks ago.

Public concern over nuclear safety is not always matched with *glasnost*. In Ukraine, where there is, not surprisingly, a high incidence of 'radiophobia', a new type of bread will soon be launched which, it is claimed, contains chemicals that will help to eliminate radioactive strontium and caesium from the body, and the appropriations for research into the effect of radioisotopes on the human organism has just been doubled by the Ukrainian Council of Ministers; developments which cause many people to doubt official assurances that the 600,000 people examined so far from the Kiev and Chernobyl areas have shown only "insignificant" levels of strontium and caesium-137. Public pressure against the building of a nuclear power station at Chyhyryn on the Dnieper resulted in its official cancellation — but, the Secretary of the Ukrainian Writers' Union, Borys Oleynyk, warned recently, construction still seems to be going on.

In Lithuania, a fire at the Ignalina RBMK power station on 20 July caused much local anxiety. The fire proved to be a relatively small one, and none of the reactors was affected, but, the station manage-

ment has been criticized by local officials and the Party daily *Tiesa* for making no announcement for five days.

And in Poland, which was also affected by Chernobyl fallout, the inhabitants of Warsaw are much concerned about a new Soviet-Polish agreement to carry out nuclear accident simulations at Swierk,

only a few kilometres from the city centre. An ageing research reactor is to be adapted for the purpose, although the scientist in charge, Dr Andrzej Strupczewski, admitted to reporters that it would probably have been easier to build a full-scale power plant. The facility, the inhabitants of Warsaw note nervously, will be only the third in the world capable of simulating a "maximum projected breakdown". The other two — one in the United States and one French, are sited respectively in the desert and underground. **Vera Rich**

# Political stalemate over IVF finally broken in Australia

## Sydney

THE political stalemate that halted activity by one of the world's top *in-vitro* fertilization (IVF) research groups has finally been broken. Alan Trounson's group at the Centre for Early Human Development, based in Melbourne's Queen Victoria Hospital, has received permission to create 80 embryos to use in tests for chromosomal defects that might occur in fertilization by the micro-injection of single sperm. Permission was granted by the Victoria state government's Standing Review and Advisory Committee on Infertility, chaired by Professor Louis Waller. The Waller committee has been deadlocked over whether to let the experiment go ahead since late 1986. The breakthrough came last month when the Victoria state parliament clarified its IVF law by finally proclaiming an outstanding section of the original legislation that allows the creation of embryos for research.

In the present experiment, the embryos must be tested within 20 hours of fertilization, just before the fusion of genetic

material from the sperm and egg occurs.

Microinjection involves placing a single sperm on the tip of an ultra-fine needle and injecting it through the outer wall of an egg. The technique offers new hope to men who are infertile due to the inability of their sperm to penetrate the egg, perhaps because of malformation of the sperm.

Research with mice has shown no increased risk of abnormalities using microinjection. In fact, several couples were provided with the microinjection treatment by a team led by Professor Carl Wood of Monash University. But the technique was then banned by the Victoria Minister for Health, pending assessment by the Waller committee. Similar techniques are already being used in the United States and Europe, although Trounson claims they have not reached his group's technical level.

The neighbouring state of New South Wales does not have any legislation corresponding to the Victoria IVF law. Trounson has collaborated with researchers there to test the microinjection technique at a private IVF clinic in Sydney, using sperm and eggs donated by 16 couples. The research was carried out under guidelines laid down by the National Health and Medical Research Council in 1982 and with the approval of the Ethics Review Committee at the Royal Prince Alfred Hospital.

The conduct of the tests in Sydney has led to calls for uniform federal regulations on IVF research. The NSW Law Reform Commission is at present examining the question of IVF and is expected to release a report shortly.

The number of embryos used in the Sydney experiments will be subtracted from the 80 to be tested by Trounson in Melbourne. He is not satisfied with the figure of 80, which was considered the minimum number that could yield meaningful results. Trounson says that the 80 will be enough to detect any dramatic increase in chromosomal abnormalities but the experiment might not pick up more subtle defects. **Charles Morgan**

# Save the trees

## Washington

IN the tradition of Live Aid and the recent 'Freedomfest', rock musician and actor Sting has joined forces with the World Wildlife Fund (WWF), the Smithsonian Institution and British schoolchildren to produce *Yanomamo*, a musical which will benefit the preservation of the world's tropical rainforests and aid the people of those areas. The first US performance was held last month at the Kennedy Center in Washington.

*Yanomamo* is a musical exploration of the problems in the Amazon Basin and a celebration of its richness and beauty. Sting hosted and narrated and the children played and sang the production. Proceeds from the event benefit WWF and the Institute for Sustainable Development, two groups devoted to conservation of tropical tracts.

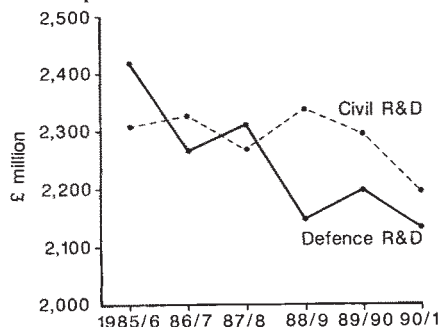
Lisa Lyles



# Gloom as Whitehall review predicts cuts in research funds

London

A BLEAK future for publicly funded UK research and development is mapped out in the annual review of government expenditure published last week. By 1990–91, the government plans to cut expenditure on civil research and development (R&D) by 6 per cent and on defence R&D by 8.5 per cent from the level of 1986–87. These cuts do not include those outlined recently for fast reactor research (see *Nature* 334, 278; 1988). Over the same period, there will be a cut in manpower of about 5 per cent. Nearly 1,500 jobs will be lost in the research councils, 726 in the Ministry of Defence and 51 in civil departments.



The downward spiral of UK R&D.

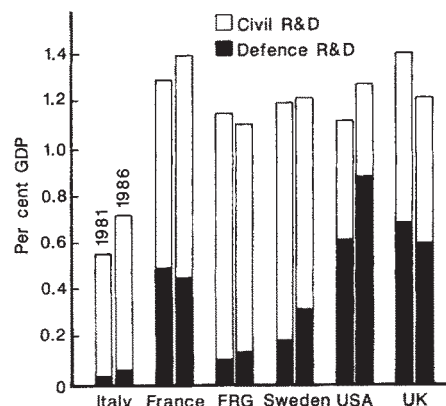
The cuts in civil R&D in part reflect government policy of withdrawing support from 'near-market' research, which it says industry should pay for. The review shows that public funds for R&D in industry have been cut drastically. But, with the exception of the chemical and pharmaceutical industry, there has been no significant increase in investment within industry. The review shows that between 1985 and 1986, R&D funded by British industry decreased from 66 to 64 per cent. In contrast, during the 1980s, foreign investment in R&D in Britain has increased substantially. From 1981 to 1986, R&D carried out in Britain by overseas-controlled enterprises more than doubled from £331.3 million to £742.3 million. During the same period, domestic industry increased R&D investment by only 55 per cent from £2,324 million to £3,621 million. But the government has until now, rejected the use of financial incentives, such as tax relief to encourage investment.

Spending on defence R&D will be £2,081 million in 1990–91, compared to £2,264 million in 1986–87. The review shows that in 1986, the proportion of government spending on defence R&D fell below the 50 per cent mark for the first time, to 49.3 per cent. But these savings are clearly not to be ploughed back into civil R&D as the science community hoped. Manpower in defence R&D will

be cut by about 3 per cent to 22,130 and in civil R&D by about 7 per cent to 20,239. Cuts are already under way in the research councils, mainly because of reductions in research commissioned by government departments. After several years of reducing staff numbers, the Natural Environment Research Council was forced this year to announce compulsory redundancies for the first time.

Government-sponsored R&D in Britain fares badly in comparison with other European countries. In 1986, Britain spent 0.62 per cent of the gross domestic product (GDP) on civil R&D, compared with West Germany (0.97 per cent), France (0.95 per cent), Sweden (0.90 per cent) and Italy (0.67 per cent). Of the countries cited in the review, only the United States (0.39 per cent) spends less in proportion to national wealth. But both the United States and Britain spend a greater proportion of GDP on defence R&D. In 1986, that proportion was 0.60 per cent in Britain compared to the United States (0.88 per cent), France (0.45 per cent), Sweden (0.31 per cent), FRG (0.13 per cent) and Italy (0.06 per cent).

The Save British Science Society calls for public investment in R&D to be linked to the national wealth. As research funds dissipate towards 1990, GDP is expected to rise by 8 per cent. The secretary of the society, Dr John Mulvey of the University



UK wealth increases but R&D lags behind.

of Oxford, says the situation is worse than most people realize. "If in 1990 we were to spend the same proportion of GDP on civil science and technology as in 1982–83, then expenditure would be £800 million greater than planned," he says.

Last week, the government advisers on the science budget, the Advisory Board for the Research Councils (ABRC), called for more money to restructure academic science and to reverse the decrease in research carried out in the research councils (see *Nature* 334, 279; 1988). Mulvey says this request is the minimum required to support the science base.

Until the Public Expenditure Survey in the autumn, the science community is hoping the Secretary of State for Education and Science, Mr Kenneth Baker, will give 'greater priority' to science, as he suggested he might in a letter to the ABRC in February. **Christine McGourty**

## NSF unveils new computer network

Washington

ACADEMIC computer users throughout the United States moved a step closer to nationwide integration last week when the National Science Foundation's upgraded computer network, NSFNET, was inaugurated. With a transmission rate of 1.5 million bits per second, and improved switching technology, the new network should allow researchers throughout the country to work interactively with NSF's six supercomputer centres as well as with databases and other resources, but demand is so great that the capacity of the network may be exhausted in as little as 18 months.

After being put out to tender last August, the contract for the upgrade and maintenance of NSFNET was awarded to Merit Inc., a consortium of eight universities in Michigan that had been successfully operating a state-wide computer network. The expansion of NSFNET by Merit has been done in cooperation with IBM, which helped to develop the switching and routing system, and MCI, a long-distance telephone company, which has supplied digital and fibre-optic land lines. Both

of the companies have provided their services free.

Computer networks in the United States have grown up through a rather haphazard aggregation of regional systems and commercial enterprises, and the resulting proliferation of electronic mail systems and transmission standards has caused as much frustration as communication. Although the expanded NSFNET unifies some of the smaller networks, it is still not easy to move between NSFNET and other national systems such as BITNET and OMNET. But according to Steven Wolf, an assistant director at NSF, plans for further unification are being studied, and NSF has expressed willingness to act as a lead agency in any such programme.

NSFNET was designed to allow an increase in transmission rate by a factor of 30, to 45 million bits per second, to be achieved as the system expands. Wolf expects that users will invent new applications as capacity and reliability increase. The most popular use today, electronic mail, was added as an afterthought to the first networks.

David Lindley

# Only the smile is left

SIR—The 30 June 1988 issue of *Nature* contains a "Scientific Paper" describing unusual results. In it, Davenas *et al.*<sup>1</sup> claim that images of molecules may be imposed on solvent such that when the original molecules are no longer present (for example, by diluting the solution sufficiently) their reactivity nevertheless persists. Specific antibodies, antigens, enzymes, ionophores and simple amines were claimed to exhibit such 'Cheshire cat'<sup>2</sup> phenomena. Because the assay used by Davenas *et al.* is closely analogous to one regularly used in our laboratory, it was simple for us to test the generalizability of their claims. We observed no results such as theirs. Evidently, whatever may have generated their own data is not readily reproduced. We therefore see no basis as yet for concluding that the chemical data accumulated over two centuries are in error, and that Davenas *et al.* have discovered a new chemical principle.

Rat basophilic leukaemia cells — a neoplastic line of mucosal mast cells of the 2H3(HR+) line<sup>3</sup> — were grown in stationary flasks and after harvesting were mixed with <sup>3</sup>H-labelled-5 hydroxytryptamine (5-HT) and mouse anti-dinitrophenyl immunoglobulin E (IgE). The mixture was distributed on 24-well polystyrene plates and incubated for 16 hours. After washing, the adherent cells were challenged with either purified rabbit anti-mouse IgE or dinitrophenylated bovine serum albumin<sup>4</sup>. In preparing sequential dilutions of the antibody and antigen, care was taken to vortex each dilution for approximately 12 seconds because Davenas *et al.* claimed this was important<sup>1</sup>. The dilutions were prepared and coded by one of us (H.M.); the assay performed and data analysed by the other (S.D.). Our results are shown in Fig. 1. No release of 5-HT was observed at high dilutions. It is apparent that the phenomenon described by Davenas *et al.* is not readily generalizable even to a closely related system.

It is reasonable to ask whether the observations of Davenas *et al.* should have been published by *Nature*. We think not. One of us (H.M.) reviewed their paper at the request of *Nature* in April 1987, and urged that the findings be checked by one or more laboratories chosen by the editor. Instead, Dr Benveniste made his own choice, and *Nature* decided to publish the report and then to despatch an international investigative team consisting of the editor, a magician and a scientist, none of whom has experience in the relevant field. Their report<sup>5</sup> provides no support for the published claims and will dismay serious scientists: it adds to the circus atmosphere engendered by the publication of the original paper. ("Homoeo-

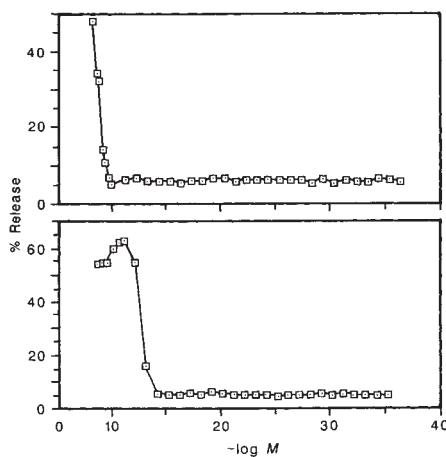
pathic enthusiasts are rejoicing while scientists scratch their heads..."; "Homoeopathy finds scientific support"<sup>7</sup>.) We believe that the approach chosen by *Nature* is regrettable. We feel that all ideas no matter how revolutionary, deserve to be heard. However, when new data are proffered that grossly conflict with vast amounts of earlier, well-documented and easily replicated data, a different editorial standard is required. Before the *imprimatur* inherent in publishing them in a leading scientific journal is granted, the new results must be reproducible by disinterested individuals familiar with the field. That is a fundamental principle of scientific objectivity. It's a shame really. It still takes a full teaspoon of sugar to sweeten our tea.

HENRY METZGER

STEPHEN C. DRESKIN

Building 10, Rm 9N258,  
NIAMS, National Institutes of Health,  
Bethesda, Maryland 20892, USA

1. Davenas, E. *et al.* *Nature* **333**, 816–818 (1988).
2. Carroll, L. *Alice in Wonderland* (Grosset and Dunlap, New York, 1987).
3. Barsumian *et al.* *Eur. J. Immunol.* **11**, 317–323 (1981).
4. Maeyama *et al.* *J. biol. Chem.* **261**, 2583–2592 (1986).
5. Maddox, J. *et al.* *Nature* **334**, 287 (1988).
6. *New Scientist* 14 July 1988 p. 39.
7. *Newsweek* 25 July, 1988 p. 66.



Release of incorporated <sup>3</sup>H-5HT from tumour basophils. Each specimen contained approximately  $5 \times 10^5$  cells and 200  $\mu$ l medium (100 mM NaCl, 5 mM KCl, 1 mM CaCl<sub>2</sub>, 0.4 mM MgCl<sub>2</sub>, 5.6 mM glucose, 25 mM Na<sub>2</sub>PIPES, 0.1% bovine serum albumin, pH 7.0). The appropriate (coded) dilution of antibody (a) or antigen (b) was added and incubation continued for 45 min. The supernatant was aspirated and counted; 250  $\mu$ l of 1% Triton X-100 in phosphate buffered saline X3 was used to release the cells and counted. Percent release was calculated as previously described<sup>4</sup>. The solid squares on each graph show the average for three buffer controls performed in duplicate for each set. All data are the means. The ranges of duplicates were  $\pm 5\%$  or less from the mean. Altogether there were 142 samples counted and all the data are shown.

SIR—In order that the controversial work of Davenas *et al.* should be considered credible, the reader must first be convinced that the data are genuine. Examination of the data represented in Table 1, however, convinces me that they are synthetic.

The results given are based on triplicate counts of approximately 80 cells and are expressed as the arithmetic mean and standard error. The standard errors vary between 0.5 and 4.8 with a median value of 1.5. This value is much less than that anticipated if normal chance variation in cell number were to have been observed. Even if no cell clumping occurred in the pipetted samples and the observers were wholly accurate, the value of the standard error is too small. In practice, this value will always be equal to or greater than the theoretical minimum standard error.

The most damning evidence is in their Fig. 1b. The number of basophils in the anti-IgG control wells, may be either greater or less than that observed in the untreated control wells, assuming that anti-IgG has no effect. It therefore follows that the "% basophil degranulation" should be less than zero in approximately 30 of the 60 dilutions used. The absence of any such 'negative' degranulation figures and the overall distribution of the results is disturbing.

As it is obvious that the data presented in Table 1 and Figure 1 are not strictly derived by experiment, the credibility of the remainder of the paper must remain in doubt.

P.M. GAYLARDE

Department of Dermatology,  
Royal Free Hospital,  
London NW3, UK

SIR—J. Benveniste and co-workers recently reported extraordinary observations on the effects of extremely diluted antibody solutions. They found that aqueous solutions of anti-immunoglobulin E (anti-IgE) still retain an ability to cause degranulation of human basophils (a type of white blood cell with cell surface antibodies to IgE) even when diluted to the point where there are no molecules of anti-IgE left in solution<sup>1</sup>. We suggest that these results are due to the production of reactive chemical species by cavitation during sample preparation.

As these authors state<sup>1</sup>, the "importance of agitation in the transmission of information was explored by pipetting dilutions up and down ten times and comparing with the usual 10-s vortexing. Although the two processes resulted in the same dilution..., degranulation did not occur at high dilution after pipetting. Ten-second vortexing was the minimum time required.... So transmission of the information depended on vigorous agitation..."

The authors are apparently unaware of



the high energy chemistry associated with turbulent flow of liquids. These chemical effects result from cavitation: the creation, growth and implosive collapse of bubbles in liquids<sup>2</sup>. The implosive collapse of bubbles can generate intense, but transient, local heating<sup>3</sup> of the order of 5,000K. This is sufficient to induce the homolytic cleavage of O-H bonds of water<sup>4</sup> and C-C bonds of hydrocarbons<sup>5</sup>. Cavitation is usually associated with ultrasonic irradiation, but it also occurs during any turbulent flow. For example, Anbar demonstrated many years ago<sup>6</sup> the existence of cavitation and associated high energy chemistry during turbulent mixing of water. The effects of cavitation and associated shock waves on biological systems can be dramatic, leading to cell rupture and death as the limiting case<sup>7</sup>.

We suggest that the degranulation observed by Benveniste and coworkers is an artefact of cell damage caused by reactions with small amounts of OH·, H·, H<sub>2</sub>O<sub>2</sub>, HO<sub>2</sub>, etc., produced by their use of vortex turbulence. Our hypothesis is easily tested: do basophils degranulate upon addition of water or buffered solutions previously subjected to vortex mixing, to high speed propeller cavitation (from a turbine homogenizer), or to high intensity ultrasound (from a cell disruptor)? The treated water should not contain any protein, which could serve to scavenge reactive species. Alternatively, do atmospheres of helium or carbon dioxide suppress the observed effects? These gases dramatically diminish the temperatures reached during cavitation collapse and suppress most chemical effects.

KENNETH S. SUSLICK

School of Chemical Sciences,  
University of Illinois,  
Urbana,  
Illinois 61821, USA

1. Davenas, E. *et al.* *Nature* **333**, 816-818 (1988).
2. Suslick, K.S., ed. *Ultrasound: Its Chemical, Physical, and Biological Effects* (VCH, New York, 1988).
3. Suslick, K.S., Hammerton, D.A. & Cline, R.E., Jr. *J. Am. chem. Soc.* **108**, 5641-5642 (1986).
4. Makino, K., Mossoba, M.M. & Riesz, P. *J. phys. Chem.* **87**, 1369-1377 (1983).
5. Suslick, K.S. & Flint, E.B. *Nature* **330**, 553-555 (1987).
6. Anbar, M. *Science* **161**, 3848 (1966).
7. Williams, A.R. *Ultrasound: Biological Effects and Potential Hazard* (Academic, New York, 1983).

SIR—Davenas *et al.*<sup>1</sup> observed repetitive waves of degranulation of human basophils, which had been exposed to increasing dilutions of anti-immunoglobulin E (anti-IgE) antibody. Degranulation was reported at the highest dilutions, even in the calculated absence of anti-IgE antibody. The hypothesis was suggested that the molecular organization of water might have been responsible for transmission of biological information. There may be a simpler explanation, more readily accommodated by the existing scientific paradigms.

Davenas *et al.*<sup>1</sup> reported using Tyrode's solution to dilute antibody. As gleaned from their paper, this solution always appeared to contain heparin. Heparin has a number of interesting properties that are germane to what may be a novel role in mediating the activity of anti-IgE antibody. Heparin exists in a helical configuration, binding water molecules, as well as a number of monovalent and divalent cations (including Na and K, which are present in Tyrode's solution)<sup>2</sup>. It also binds to various proteins and synthetic polypeptides<sup>3</sup>. Like other glycosaminoglycans, heparin chains are further capable of interacting with one another, forming molecular aggregates<sup>3</sup>.

I propose that anti-IgE antibody (or any of the other immunological stimuli noted in the paper<sup>1</sup>, that were responsible for basophil degranulation) might have acted as a template for heparin, thereby inducing a specific conformation of the heparin molecule. This molecular conformation would then be stabilized, perhaps by interacting with another heparin molecule. Upon dilution with heparin-containing Tyrode's solution, the stabilized heparin conformation, although lacking biological activity, would itself serve as a template, effecting a new heparin conformation which would mimic the three-dimensional structure of the antigen-binding site of anti-IgE antibody (or other immunological stimulus). Presumably water molecules and perhaps the cations present in Tyrode's solution would stabilize this new, antigen-binding heparin conformation. Subsequent dilution with fresh heparin would, however, result in the antigen-binding conformation further acting as a template for formation of the biologically inactive conformation. Successive dilutions would generate the alternating heparin conformations.

J. LESLIE GLICK

Bionix Corporation,  
10899 Deborah Drive,  
Potomac,  
Maryland 20854, USA

1. Davenas, E. *et al.* *Nature* **333**, 816-818 (1988).
2. Chakrabarti, B. & Park, J.W. *Crit. Rev. Biochem.* **8**, 225-313 (1980).
3. Lindahl, U. & Höök, M. A. *Rev. Biochem.* **47**, 385-417 (1978).

SIR—There may be a very simple explanation of Benveniste's conclusion (*Nature* **333**, 816; 1988). Imagine there is some degranulating active molecule, not necessarily an anti-IgE, that binds to one component of Hepes-Tyrode's solution, for example heparin or EDTA; then there will not be a real dilution of the active compound.

Imagine that the binding is or is not reversible depending on the degree of "vigorous shaking", you get an explanation for "rhythmic fluctuation" in the activity (not so rhythmic in fact). This

explanation is more comfortable than 'water memory' or throwing away the Law of Mass Action or Avogadro's number.

M.J. ESCRIBANO

CNRS Groupe de Laboratoires de  
l'Institut de Recherches Scientifiques sur  
le Cancer,  
7 Rue Guy-Môquet,  
BP No. 8,  
94802 Villejuif Cedex,  
France

## Creationism and evolution

SIR—Andrew P. Whipple (*Nature* **333**, 492; 1988) tries to take an even-handed approach towards science and creationism, but in my opinion fails in this attempt. First, he fails to give weight to the coherence between theory and observation that evolutionary theory is alone able to offer. It is this observationally multicentred coherence that gives scientific 'belief' a privileged character, not shared by other types of belief. Furthermore, Whipple gives a truncated account of the 'world views' in presence. He defines the naturalistic viewpoint as one of two basic positions, namely one that asserts that there is no reality beyond the physical and that denies the supernatural. This he opposes to the theistic viewpoint. In fact, there is a third position, not covered by your correspondent's analysis, which represents a different naturalistic view. This view also excludes the supernatural, but holds that the spiritual domain has a reality of its own as much as the world of phenomena does, a reality subject to the equivalent of natural laws, that is, engaged in obligatory, if unknown, relationships with the world of phenomena. These relationships need not be more arbitrary or inconstant than the relationships within the observable world itself and therefore are not in favour of a personalized and humanized god. The view I refer to holds that the mind, fully as real as, and distinct from, observable nature, is another side of nature. (By observable nature, I refer to the nature before us, not necessarily totally observable.) Our present ignorance of what this other side of nature is does not dispense us from acknowledging the existence of a philosophical position that is opposed by creationism, yet is not 'materialistic', but 'naturalistic' in a sense different from your correspondent's.

EMILE ZUCKERKANDL

Linus Pauling Institute  
of Science and Medicine,  
440 Page Mill Road,  
Palo Alto,  
California 94306, USA

Letters submitted for Correspondence should be typed and double-spaced.

# What price academic freedom?

*The British government now has its Education Reform Act. Academics and their leaders should stay alert until the government's intended uses of its powers are made clear.*

ACADEMICS, like farmers, are an ungrateful and complaining lot. That, at least, must be the opinion of the British government, which last week carried its Education Reform Bill through the British parliament more or less intact. For the grumbling continues among academics even though the government has conceded on a number of important if symbolic issues — "academic freedom" has a legal definition, the Universities Funding Council will be able to offer advice to the government as well as to obey such directions as are handed down and, more important, these directions will be, in principle at least, open to scrutiny and debate. Why, with all this conceded, are academics still grumbling?

The concessions are significant, and a considerable achievement for the Committee of Vice-Chancellors and Principals, which organized the campaign against the bill. The reason why discontent persists is mostly that the Education Reform Bill has brought about such a radical change of the rules that British academic life will never be the same again. That it should be necessary to define academic freedom in a legal statute gives the show away. The essence of the change that has been brought about is that British universities have ceased to be independent autonomous self-governing institutions, but have become dependent on the state not merely for funds but for direction. Many of the new-fangled instruments of public administration which the present government has recently devised, known as 'trading funds' or 'agencies', will enjoy more freedom than the run-of-the-mill British university.

It is not far-fetched to compare what has happened in Britain with what happened to the University of Malta, then the oldest university in the British Commonwealth, in the 1970s. The then Prime Minister of Malta, Mr Dom Mintoff, got it into his head that the university, by being selective and concerned with academic objectives, was elitist and unproductive. In such a small community, it was therefore natural that the university should be required to change its way of doing business. Knowing they had no right to resist, but only that of contrary persuasion, the university authorities fell in with Mintoff's more moderate proposals, weakening still further their capacity to mount effective resistance against his more malign intentions. By 1978, the university found certain of its

courses being transferred to Malta's only other institution of higher education, the polytechnic in Valetta. Soon afterwards, the university was defunct.

While Mintoff and the present British government are poles apart on politics, their opinions of universities have one crucial element in common: each considers that universities by effeteism have failed society. The government of Malta in Mintoff's time believed that its single university had failed to propagate egalitarianism, the British government believes its larger university system should have made Britain more prosperous much sooner. By decreeing in the 1970s that the government of Malta knew best how and where the university's courses should be taught, Mintoff was not necessarily going much further than the British government's plan that the Universities Funding Council should issue directions to universities in general and in particular about the balance of their teaching effort.

Some will protest that university systems as dependent on the state for funds as the British are always vulnerable, but that is an exaggeration. On mainland Europe, all university systems are state-financed and even state-governed, but academics are not as much threatened as under the Education Reform Act, as it has become. In West Germany, the responsibility of the *Länder* governments for higher education means that universities are at the centre of regional pride and aspiration. In France, despite the status of most academics as public employees, there is enough general respect for the independence of the academic life for the ground rules to have been virtually unchanged for decades, the upheavals of 1968 notwithstanding.

Britain's boast of the benefits of its distinctive means of laundering public money intended for university support now seems distinctly hollow. Originally, there was a University Grants Commission dependent directly on the Treasury which channelled funds towards the universities, largely on its own responsibility. Then tidy-mindedness suggests that the commission should become a committee, dependent on the Department of Education. That was in the early 1960s, a period of rapid expansion in British higher education buoyed up by economic optimism. Then followed the longer period in which ambitions were lowered and budgets were constrained, matched meticulously

against the needs of schools and of vocational education. The universities would have been better able to defend themselves against the events of this decade if they had not squandered so many of their political friendships in the previous decade, but they had no obvious reason to fear that the present government's animus against academic life would declare itself in the arbitrary 10 per cent plus cut in university budgets decreed at the end of 1980.

British and other academics know too well what has happened to British universities since then. As at the University of Malta in the late 1970s, the system has been weakened and demoralized to the point at which a push will send it reeling. What tends to be forgotten is that the much smaller university system immediately after the Second World War was in many ways much stronger than its successors have now become. Nostalgia is a poor guide to action now, because the circumstances are different, but the past should not entirely be forgotten.

Take, for example, the University of Manchester in the early 1950s, with a student body half its present size. Within a few years of the Second World War, this institution, already distinguished for its historical studies by the presence of Lewis Namier, had built an innovative radio-astronomy observatory (Jodrell Bank), had discovered strange particles in cosmic rays, had founded palaeomagnetism and built a digital computer that could have been the foundation of an industry now flourishing mostly elsewhere. Yet Manchester, if unusual, was by no means unique. And academics were broadly speaking as free as the air. They rarely complained that they had to work like dogs, but they were paid appallingly.

No purpose can be served by hoping that those happy times will spontaneously return, or that they would seem happy if they did. Things have changed. The participation rate in higher education has multiplied, in Britain, three times, and there is every reason for hoping that, in the national interest, it will become larger still. The best hope is that the government means what its spokesmen have been saying during the debates on the Education Reform Act, that it will not use the potentially malevolent powers the law now gives it. The vice-chancellors had better not disband their lobby until that is clear.

**John Maddox**



## Cognitive neuropsychology

## Sensation and semantics

John C. Marshall

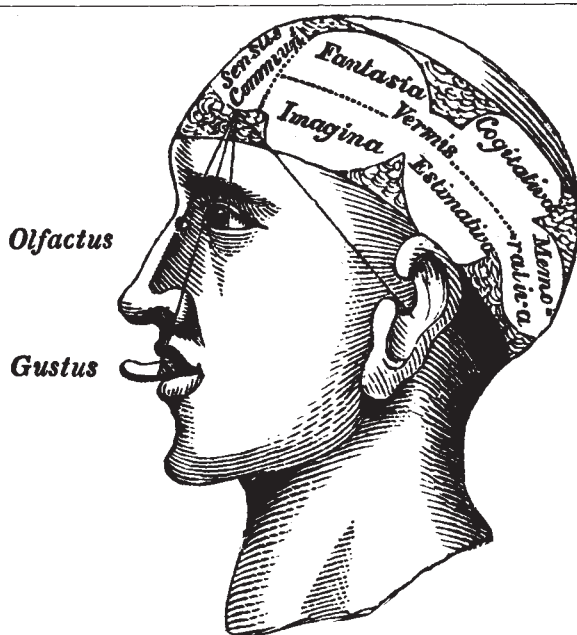
"*NIHIL est in intellectu quod non primum in sensu*" (Nothing is understood by the intellect which is not first perceived by the senses) — thus wrote Aristotle some 2,000 years ago. The motto has stood the test of time pretty well, despite some chipping away by those who argue that the innate organization of the central nervous system imposes severe constraints upon the mind's interpretation of sensation<sup>1,2</sup>.

Nonetheless, it would be quite rational to suppose that although thoughts must be triggered by experience, the conceptual and encyclopaedic knowledge so accessed is coded in a form independent of sensory modality. Suppose I believe pigs to be "unclean, medium-sized quasi-domestic animals providing pork and bacon for the table". Surely I access that neurologically unitary knowledge, irrespective of whether I see, smell or hear a pig, or see or hear the word pig? To suggest duplication of such knowledge — that is, an identical or at least highly similar body of information is stored once in conjunction with the look of pigs and once again with the sound of pig — seems decidedly counter to common sense. And yet, if the interpretation that McCarthy and Warrington<sup>3</sup> place upon their results, reported on page 428 of this issue, is correct, common sense would appear to be wrong.

It is now reasonably well established that acquired brain damage can selectively impair knowledge within specific conceptual categories<sup>4</sup>. McCarthy and Warrington's patient (T.O.B.) sustained relatively focal damage to the left temporal lobe that resulted in a selective inability to define the characteristics of animals or plants in response to their spoken names; by contrast, he is consistently much better at providing information about inanimate objects. Thus in response to 'pig', T.O.B. could only volunteer that it was an animal, whereas 'wheelbarrow' elicited the reply, "the item we have here in the gardens for carrying bits and pieces; wheelbarrows are used by people doing maintenance here on your buildings. They can put their cement and all sorts of things in it to carry it around". So far so good: comparable phenomena of category-specific impairment have now been widely reported in cases of both acquired<sup>5</sup> and developmental<sup>6</sup> pathology.

But the performance of T.O.B. adds a

further puzzle to the pattern. Although he is unable to define adequately the spoken names of living things, T.O.B. succeeded in providing the relevant details when given a picture to identify. In response to the word 'rhinoceros', he can only manage, "Animal, can't give you any functions"; yet a picture thereof elicits the information "Enormous, weighs over one ton, lives in Africa". From this reliable



Our expression 'common sense' arises from Aristotle's theory that there is a general sense common to all the special senses. The communis sensus is seen at the front of the brain in this illustration, first published in a translation of J. F. Blumenbach's *Institutiones physiologicae*, published in 1840. This is, in fact, the location of the frontal cortex, now known to be involved in integrating information from the various senses. (From ref. 9.)

pattern of performance, McCarthy and Warrington draw the conclusion that semantic information is multiply represented in the normal brain and that this duplication is linked to the input modalities whereby we gain knowledge of the world. (For purposes of the argument, language must be viewed as a discrete input system, analogous in this regard to the senses<sup>7</sup>.) The conclusion that T.O.B. has sustained modality-specific loss of information about living things, rather than differential impairment of access to such knowledge, is reinforced by the observation that he is highly consistent with respect to the particular items on which he fails.

McCarthy and Warrington support their hypothesis with an ontogenetic argument. Let us suppose, not unreasonably, that I first acquired the knowledge that whales are large sea-living creatures

in conjunction with pictures of whales; but that I later learned verbally that 'whales', by virtue of their biology, are mammals not fish. This developmental history could accordingly programme two neuronal substrates, in the visual association cortex and auditory association cortex, with different pieces of information about the same referent. If there was mutual transfer of information between these loci, duplicate encyclopaedias could develop, each accessed by a particular sensory modality. In principle, acquired brain damage could thus destroy one information store whilst leaving the other relatively intact. The upshot would be a modality-specific loss of knowledge, exactly the result that McCarthy and Warrington report<sup>3</sup>.

A similar dissociation could occur without duplication of information. Suppose that knowledge acquired in conjunction with a picture was stored uniquely in visual association cortex, and that knowledge acquired in conjunction with the name of the picture's referent was stored only in language-committed cortex. Bidirectional associative links between the two sets of information could then enable the normal brain to retrieve the full corpus of knowledge, irrespective of whether the system is addressed through picture or word. But if these linkages were broken, presentation of word or picture would only access the knowledge originally acquired in each respective condition.

On this hypothesis, the brain could store contradictory information about the same referent without consciously bringing the anomaly to mind. This might seem like a *reductio ad absurdum* of the notion of modality-specific cognitive deficits. But before we dismiss

the idea, we should recall a patient (A.R.) reported by Warrington and Shallice<sup>8</sup>. To the auditory stimulus 'Solzhenitzyn', A.R. responded that this was the name of a living Russian novelist. But the written stimulus *Solzhenitzyn* he claimed was the name of a Polish novelist who died about 1900. Our understanding of the brain still lies in the heart of darkness. □

1. Leibniz, G.W. *New Essays on Human Understanding* (Cambridge University Press, 1981).
2. Chomsky, N. *Knowledge of Language* (Praeger, New York, 1986).
3. McCarthy, R. A. & Warrington, E. K. *Nature* **334**, 428–430 (1988).
4. Shallice, T. *Cognitive Neuropsychol.* **5**, 133–142 (1988).
5. Warrington, E. K. & Shallice, T. *Brain* **107**, 829–854 (1984).
6. Temple, C.M. *Brain* **109**, 1225–1242 (1986).
7. Marshall, J.C. *Cognition* **17**, 209–242 (1984).
8. Warrington, E. K. & Shallice, T. *Brain* **102**, 43–63 (1979).
9. Clarke, E. & Dewhurst, K. *An Illustrated History of Brain Function* (Sandford, Oxford, 1972).

John C. Marshall is in the Neuropsychology Unit, Department of Clinical Neurology, Radcliffe Infirmary, Oxford OX2 6HE, UK.

## Surface science

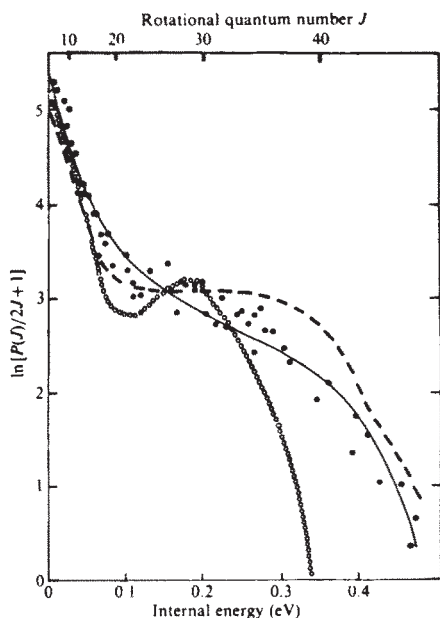
## New directions in gas scattering

David A. King

STUDIES of the scattering of molecules from surfaces usually cannot distinguish the different orientations of the molecule at the moment of impact. But intuition and models suggest that the orientation of asymmetric molecules could be crucial. On page 420 of this issue<sup>1</sup>, E. W. Kuipers *et al.* from the Atomic and Molecular Physics Institute in Amsterdam demonstrate experimentally that this is the case. They show that rotationless nitric oxide molecules gain more rotation when scattered off a silver surface if they are oriented oxygen-end forwards than with the nitrogen-end first.

This work represents the most recent advance in extending purely kinetic measurements of gas-solid interactions to dynamic studies using molecules in selected quantum states. This challenging experimental area combines state-of-the-art laser technologies developed in studies of gas-phase chemical reaction dynamics with surface-science techniques. Indeed, the most successful groups in the field bring together experts from the two disciplines.

Supersonic molecular beams are the *sine qua non* of these studies. Molecules at



**Fig. 1** Boltzmann plot of the final-state rotational distribution of NO molecules scattered from Ag(111) (filled circles)<sup>3</sup>. The ordinate is chosen so that if the rotational distribution were in thermal equilibrium, it would be plotted as a straight line with a negative slope on the graph. The data are compared with a quantum-mechanical calculation (open circles), a classical Langevin simulation (dashed curve) and a phenomenological analysis (filled circles and solid curve). (From ref. 4.)

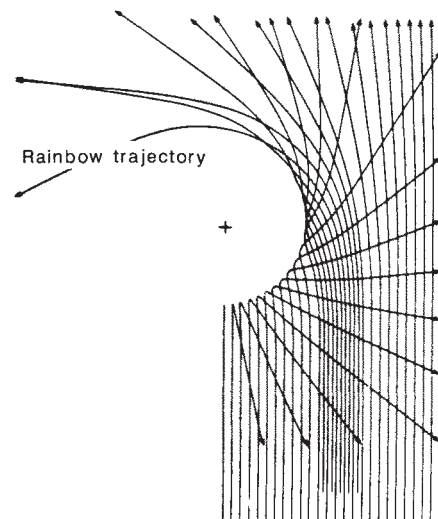
high pressure are squeezed through a pin-hole into a vacuum. Their short collisional mean-free path within the beam means that they efficiently approach thermodynamic equilibrium with most of their thermal energy going into forward motion. This produces a large flux of vibrationally and rotationally cold molecules with a narrow spread of translational energies so that the collision energies are well known.

Researchers have concentrated on determining the translational and rotational energy distributions of molecules scattered from a clean, single crystal plane of a solid surface. They use a combination of mass spectrometric time-of-flight analysis, laser-induced fluorescence (LIF) and multiphoton ionization (MPI) to detect the scattered molecules. The scattering of nitric oxide (NO) from the silver (111) surface has become the prototype for the rotational studies<sup>2</sup>, primarily because the appropriate excited states of NO are readily accessed by lasers in the visible part of the spectrum.

A 'Boltzmann' plot (Fig. 1) of the rotational distribution (as a function of  $J$ , the rotational quantum number) of NO molecules scattered from Ag(111) would be a straight line of slope  $-1/kT_f$ , where  $T_f$  is the rotational temperature and  $k$  is Boltzmann's constant, if the final distribution were thermal (statistical). But this behaviour is observed only at low final-state energy ( $J < 20$ ) and the final-state temperature matches that of neither the crystal nor the incident beam<sup>3,4</sup>.

Figure 1 also demonstrates the input from various theoretical models. Quantum-scattering theory (open circles) includes an anisotropic term which accounts for different behaviour for the two ends of the NO molecule. The results show that the high- $J$  plateau in the data of Fig. 1 is due to a rotational rainbow: the superposition of many collisions leading to similar rotational excitation (Fig. 2). A classical calculation (dashed line) also attributes the plateau to rainbow scattering.

In these experiments it seems that thermal averaging swamps quantum effects. More state selection is needed before a detailed understanding of the dynamics of the interaction can be achieved. For example, optical pumping using stimulated emission, Raman pumping or infrared excitation can be used to prepare single vibrational states. Further developments can also be anticipated in the analyses of scattered or desorbed molecules, particularly using resonance-enhanced multiphoton ionization (REMPI), which will yield both angular



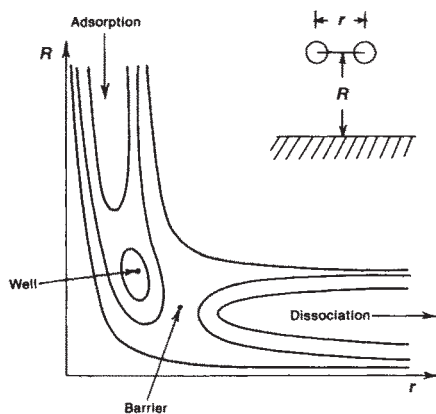
**Fig. 2** Rotational rainbows are named by analogy with classical rainbow scattering of single particles off a point potential as shown above. Trajectories either side of the rainbow trajectory give rise to less extreme scattering. Paths close to the rainbow trajectory build up to give strong scattering in the direction of the rainbow angle. Similarly, scattering trajectories of diatomic molecules converge to an extremum that gives maximum rotational excitation after the collision, a phenomenon known as the rotational rainbow. (From ref. 4.)

and state resolution.

The anisotropy of the interaction of NO with the silver surface is confirmed in the new work of Kuipers *et al.*<sup>1</sup>. They use a supersonic source to generate a molecular beam of NO at such low rotational temperatures that effectively all molecules are in the rotational ground state. The molecules are then aligned by a strong electric field before striking the silver (111) surface. By switching the polarity of the field, the molecular alignment can then be changed to favour either the N or the O end of the molecule to collide with the surface. The molecules scattered in the specular direction are spatially analysed and show that O-end collisions produce significantly more rotational excitation in the scattered NO molecule than result from N-end collisions.

Studies of the sticking of molecules onto a surface, their dissociation into atoms and their desorption could also be enhanced by improved state selection. These are of particular interest for elucidating the details of catalysis. Since the pioneering work of Ballooch *et al.*<sup>5</sup> on the dissociative sticking probability of H<sub>2</sub> on copper as a function of the incident collision energy, several studies have been performed on systems with and without an activation barrier to adsorption (Fig. 3). In the former case, the probability of dissociative adsorption increases as the normal component of the incident kinetic energy exceeds the barrier height; in the latter case, the sticking probability generally decreases with increasing kinetic energy. The vibrational state of the incident beam





**Fig. 3** Contour plot of the potential energy of interaction between a molecule and a surface. It can be visualized as depicting two valleys running parallel to the axes with steep walls rising to the axes and a high central plateau. As the molecule approaches the surface ( $R$  decreases) its energy rises towards the potential barrier. If the energy can be reduced by letting the two atoms of the molecule separate on the surface ( $r$  increases), the molecule dissociates. A well at small  $r$  and  $R$  represents the molecule bound whole to the surface. The effect of molecular orientation is not shown. Whether the potential energy maximum occurs as the molecule approaches the surface or as it dissociates can have a crucial effect on the reaction dynamics of the system<sup>6</sup>. (From ref. 7.)

may also influence the probability of sticking. Pursuing the investigation of Kuipers *et al.*<sup>1</sup>, it would be interesting to determine whether or not the orientation of the incident molecule has any observable effect on sticking.

Theoretical work in this area has kept pace with experimental results. Calculations of the adsorption potential energy surface are complicated by the fact that even for a diatomic molecule there are six degrees of freedom: the orientation and rotation of the molecule, its bond length and vibrational state, and its separation from the surface and velocity. But useful calculations have been made by fixing several of these for particular systems —  $H_2$  on  $MgO$  and  $Ni$ ;  $N_2$  on  $Fe$ . Trajectory calculations based on these surfaces give useful insights into the problem. In gas-phase chemical reactions, whether the activation barrier occurs as the reagents approach or as the products separate can have a critical effect on the reaction dynamics. The same effect should be observed in gas-surface collisions<sup>6</sup>. □

1. Kuipers, E. W., Tenner, M. G., Kleyn, A. W. & Stolte, S. *Nature* **334**, 420–422 (1988).
2. Barker, J. A. & Auerbach, D. J. *Surface Sci. Rep.* **4**, 1–100 (1984).
3. Kleyn, A. E., Luniz, A. & Auerbach, D. *Phys. Rev. Lett.* **47**, 1169–1172 (1981).
4. Zangwill, A. *Physics at Surfaces* (Cambridge University Press, 1988).
5. Ballooch, M., Cardillo, M. J., Miller, D. R. & Stickney, R. E. *Surface Sci.* **46**, 358–370 (1974).
6. Ertl, G. *Ber. Bunsen Phys. Chemie* **86**, 425–432 (1982).
7. Levine, R. D. & Bernstein, R. B. *Molecular Reaction Dynamics and Chemical Reactivity* (Oxford University Press, 1987).

David A. King is in the Department of Physical Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EP, UK.

## Mantle structure

# Layer cake or plum pudding?

Richard W. Carlson

Is the Earth's mantle chemically stratified or are there chemically distinct regions of various sizes distributed like plums in a pudding throughout the whole mantle? Answers to this question, addressed at a recent meeting<sup>\*</sup>, are important not only for the obvious implications for the present structure of the mantle but also because chemical stratification could control the nature of solid-state convection in the Earth's interior. If the compositionally induced density contrasts between layers are sufficiently large, convection will be confined within the individual layers with little or no mass transport between them.

The upper mantle is richer in the mineral olivine than would be expected if the Earth had formed by accumulation of chondritic meteorites. Because of the high compressibility of silicate liquids relative to that of solids, increasing pressure in the Earth causes silicate liquids to change from being less dense than olivine to being more dense at a depth of about 250 km. If the outer part of the Earth melted extensively during formation, crystallization of olivine with its subsequent sinking and floating towards 250 km depth could result in a chemically stratified Earth with an olivine-rich upper mantle (D. Walker, Lamont-Doherty Geological Observatory).

For models that start with an extensively molten Earth, it is necessary that minerals such as majorite and perovskite, which can crystallize from the magma at great depth, do not fractionate elements in a way that produces upper-mantle compositions different from those observed. The first measurements of elemental distribution coefficients for majorite and perovskite (A. E. Ringwood, Australian National University) show, however, that fractionation of these minerals will cause the residual liquid to move towards compositions unlike those observed for the upper mantle. If confirmed, these results suggest that these high-pressure phases have not participated in a significant way in crystal-liquid fractionation as the Earth evolved. By inference, this implies that the Earth may never have been molten at depths where these phases are stable.

The elemental abundance pattern of the upper mantle suggests that it has had some partial melt removed from it but the melt seems to be similar to the basaltic magmas produced by melting at shallow (less than 150 km) depths. The volume of lost basalt necessary to explain the composition of

the upper mantle, assuming that the upper mantle is of constant composition, is roughly 2.5 times the mass of the existing continental and oceanic crust combined. Consequently, there must be a reservoir of this basaltic material in the Earth's interior. One possibility is that the basalt fraction removed from the upper mantle has been transferred and mixed into the lower mantle through the subduction of oceanic plates (E. Takahashi, Okayama University; my own work).

But new measurements (Ringwood) indicate that when basaltic oceanic crust is reinjected into the mantle, it is more dense than most upper-mantle materials but less dense than the perovskite mineral phases that constitute the lower mantle. This could cause subducted crust to accumulate at the upper-lower mantle boundary. Support for this suggestion may come from the distinct seismic properties of the 400–650-km-deep transition zone between the upper and lower mantle (D. L. Anderson, California Institute of Technology). But the current level of uncertainty in the laboratory data for the seismic properties of materials at these pressures and temperatures allows, or even favours, a transition zone composed of the same material as the remainder of the upper mantle (D. J. Weidner, State University of New York, Stony Brook).

The retention of subducted oceanic crust in the upper mantle contradicts seismic evidence that subducted plates penetrate through the 650-km-deep upper-lower mantle boundary (T. H. Jordan, Massachusetts Institute of Technology) and perhaps continue all the way to the core-mantle boundary (A. M. Dziewonski, Harvard University). Penetration of subducted plates to the core-mantle boundary would mean that the upper and lower mantle are not separated, although the Earth could still be described as layered. If the chemically induced difference in density between the upper and lower mantle is about 2 per cent, which is consistent with recent comparisons between mineral physics and seismic data for the lower mantle (H. Mao, Carnegie Institution of Washington; W. A. Bassett, Cornell University), the thermally induced high density of the cold subducting plates should allow them to sink through the 650-km boundary without entraining much upper-mantle material.

When these plates become warmed by the lower mantle, they rise again into the upper mantle without carrying much lower mantle with them. This form of penetrative convection (P. Silver, Carnegie

<sup>\*</sup> Twice-yearly meeting of the American Geophysical Union, Baltimore, 16–20 May 1988.

Institution of Washington; P. Olson, Johns Hopkins University) allows the subducted plates to remove heat from the lower mantle but results in little net mass transfer between upper and lower mantle.

The return of subducted plates into the upper mantle is expected to be concentrated in areas of convective upwelling in the lower mantle. These areas, identified by their relatively slow seismic velocities on global tomographic maps of the lower

mantle (Dziewonski) do seem to coincide with areas where young oceanic basalts have anomalous isotope compositions (P.R. Castillo, Carnegie Institution of Washington) suggesting the first direct correlation between features of the lower mantle and the geochemical properties of the Earth's surface. □

*Richard W. Carlson is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA.*

## Protein structure

# What's left out tells the story

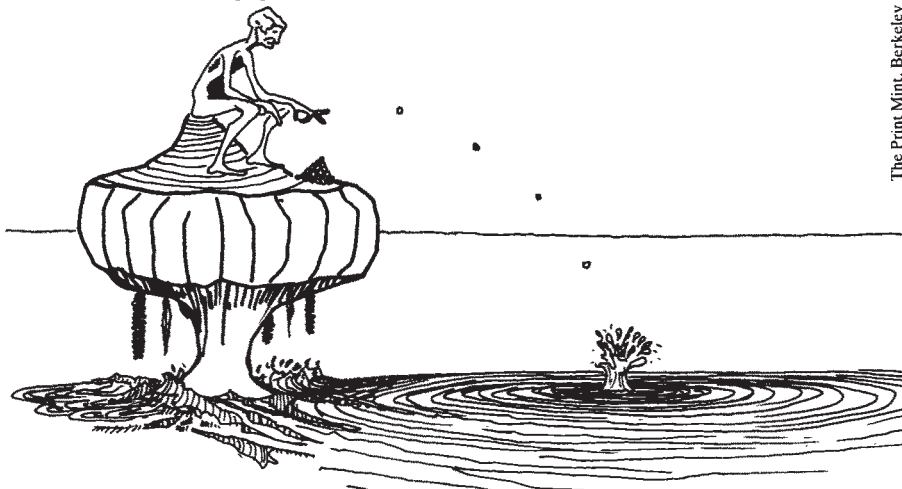
C.E. Schutt

"RUBIK's nightmare" is a three-dimensional jigsaw puzzle having hundreds of irregular, slightly sticky, solid pieces strung out at precise intervals along a semi-flexible chain. For some, it is a two-person underwater game, where one player tries to disassemble what the other is piecing together<sup>1</sup>. The essence of the problem is that the chain cannot acquire strain as the pieces snap neatly into place without leaving interior voids, much like those neat little packages we call proteins. Two recent papers, one by Alan Fersht and colleagues<sup>2</sup> and the other by Brian Matthews and colleagues on page 406 of this issue<sup>3</sup> chip away at the solution, methyl group by methyl group, by providing direct proof that introducing cavities into proteins leads to less stable structures.

Pauling deduced from the planarity of the peptide bond and the precise geometry of hydrogen-bond donor-acceptor pairs that backbones of proteins could form regular, unstrained helices and sheets despite the chemical diversity of amino-acid side chains — "the law of the chain". Left unexamined, our human tendency would be to read into the appearance of such regularity an underlying basis for stability. In fact, short isolated  $\alpha$ -helices and  $\beta$ -sheets are, with few exceptions, not very stable because solvent water molecules can form hydrogen bonds with the backbone with equal ease<sup>4</sup>. What stabilizes globular proteins in aqueous solution is the net entropy gained as aliphatic side chains are buried in the hydrophobic core<sup>5</sup>. Completing the "law of side chains" is F.M. Richards' canonical dictum, "eschew voids", a compact summary of observations made on many proteins that interior hydrophobic residues are nearly close packed.

Like Newton's laws it is not at once obvious how one should apply these principles to solve specific molecular-design problems which involve, among others, tradeoffs between backbone distortions and solvent entropy. What has hampered attempts to develop a precise analytical

theory of protein stability is the very small free-energy difference between the native and unfolded states, something of the order of only 10 kcal mole<sup>-1</sup> (ref. 6). From the perspective of our main-chain-side chain balancing act this represents approximately the energy required to isomerize just a single peptide bond<sup>7</sup>, or to



*An epitome of the protein stability problem. Exposure of methyl groups, represented as small stones, to solvent leads to instability of the whole structure. Eventually the seated crystallographer is undermined. From an original poster by Hey Mack.*

create a void by exposing a single large hydrophobic group to solvents. A useful notion is that of solvent-accessible surface area, which prescribes a free-energy gain of approximately 20 cal mole<sup>-1</sup> for each square angstrom of buried hydrocarbon surface<sup>8</sup>. Although many tests of this concept have been proposed, including a bold application to a *de novo* protein, each in some way relies on indirect correlation with measurements on the free energy of transfer of amino acids from organic to aqueous solvents<sup>9</sup>.

The new work of Fersht's<sup>2</sup> and Matthews's<sup>3</sup> groups provides some data that bear on the problem. In the experiments of both groups, single-site mutations were made at known positions in proteins resolved at high resolutions so that cavity geometry could be calculated. These values are compared directly with

measurements of free energy obtained by thermal or urea denaturation. In the case of barnase, the protein alterations are as small as possible, removal of just a single methyl group, introducing only the merest complications into the interpretation of the urea denaturation experiments. This is an extremely important point that has been discussed recently in the News and Views article<sup>10</sup> by Walter Kauzmann. (For lysozyme, the changed residue is part of the amino-terminal surface-bound chain.) Given that the chain may play according to its own rules, the mutants were subjected to high-resolution X-ray crystallography to verify residue positions. It is intriguing that the only buried mutant residue (Ile → Val) that crystallizes is the same truncated, or minimal, mutation studied for barnase. Clearly, more analysis of molecular contacts in the crystal are required and perhaps even crystal-melting data will be needed to balance the 'nooks' against the 'crooks'.

Finally, paraphrasing Hesse's *Magister Ludi*, perhaps our "bead game" is moving away from the chain-based folding

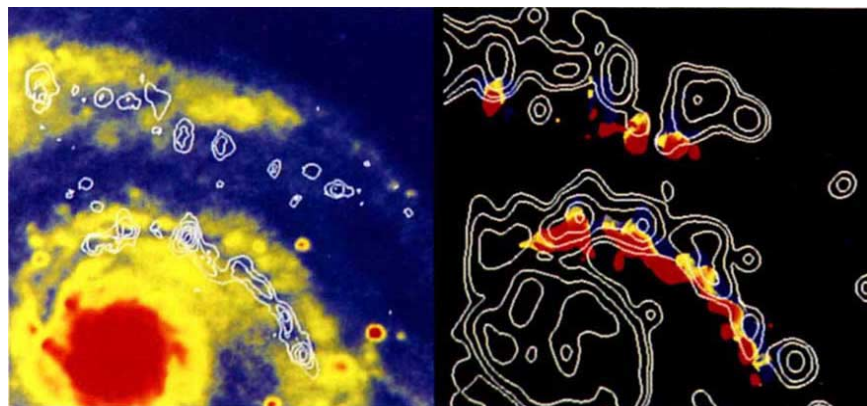
problem, maybe only an epiphenomenon, to anticipating all possible side-chain packing combinations and an answer to the question of why nature needs three isochothic side chains<sup>11</sup>. □

1. Harrison, S.C. & Durbin, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4028–4030 (1985).
2. Kellis Jr, J.T., Nyberg, K., Šali, D. & Fersht, A.R. *Nature* **333**, 784–786 (1988).
3. Matsumura, M., Becktel, W.J. & Matthews, B.W. *Nature* **334**, 406–410 (1988).
4. Shoemaker, K.R. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2349–2353 (1985).
5. Kauzmann, W. *Adv. Prot. Chem.* **14**, 1–63 (1959).
6. Pace, C.N. *Meth. Enzym.* **131**, 266–279 (1986).
7. Winkler, F.K. & Dunitz, J.D. *J. molec. Biol.* **59**, 169–182 (1971).
8. Lee, B. & Richards, F.M. *J. molec. Biol.* **55**, 379 (1971).
9. Eisenberg, D. & McLachlan, A.D. *Nature* **319**, 199 (1986).
10. Kauzmann, W. *Nature* **325**, 763–764 (1987).
11. Ponder, J.W. & Richards, F.M. *J. molec. Biol.* **193**, 775–791 (1987).

*C.E. Schutt is in the Department of Chemistry at Princeton University, Princeton, New Jersey 08544, USA.*



## Riding along on the crest of a wave



THE familiar spiral structure of many galaxies has more than visual appeal: gas and dust, the debris of stellar evolution, accumulate in the spiral arms to form a new generation of stars. But how spiral structure forms, and how it influences or perhaps controls star formation is not clear.

In the theory of density waves, spiral arms result from collective gravitational motions in the galactic disk, and make a travelling wave that orbits the disk. But not all galaxies have the perfect curved spokes of 'grand design' spirals. Often the arms are fragmented, leading some astrophysicists to suggest that they are purely local phenomena, perhaps the result of supernovae blowing out shells of material which get stretched into arcs by the differential rotation of the galaxy. Star formation happens as the supernova shock compresses existing clouds of gas, and the spiral arms are a consequence, not a cause, of star formation.

On page 402 of this issue, S.N. Vogel *et al.* present observations which favour the density-wave theory. On the left, contours representing the intensity of CO emission are superposed on an infrared image of the galaxy M51, showing that dense molecular clouds coincide with dust lanes in the spiral arms.

On the right, the colours represent the line-of-sight velocity of the CO emission relative to the velocity of the galaxy (red means motion towards us, blue away) and show a sharp change in velocity as the gas moves across the spiral arm. The contours are of hydrogen emission, associated with young stars; stars seem to be forming just downstream of gas that has passed through the density wave. These observations supply a consistent picture of molecular clouds being triggered into star formation as the density wave passes. □

## Cosmic rays

# More missing neutrinos

A.W. Wolfendale

PROTONS may be reticent about decaying in public, but the detectors built to search for their decays are providing fascinating science. Coming hard on the heels of the first detection of neutrinos from supernova 1987A in the Large Magellanic Cloud in February last year, the Kamiokande group in Japan now reports a result which, if true, has an interpretation more important still. Hirata *et al.*<sup>1</sup> claim that their 3,000-ton detector located 1,000 m underground has detected too few muons produced by neutrinos interacting inside it, these neutrinos having been produced in the atmosphere by incoming cosmic-ray particles. Electrons produced in the detector by a similar process seem to appear at the expected rate, however.

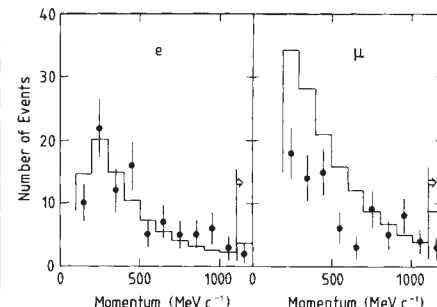
As the authors conclude: "some as-yet-unaccounted-for physics such as neutrino oscillations might explain the data". The Holy Grail of neutrino physicists — neutrino oscillations — might be at hand.

The neutrino ('the little neutral one') was postulated by Pauli to solve the problem of the missing energy in nuclear  $\beta$ -decay, so that the basic reaction producing  $\beta$ -particles (electrons) became neutron  $\rightarrow$  proton + electron + neutrino or, more precisely,  $n \rightarrow p + e^- + \bar{\nu}_e$ , the bar indicating that the neutrino is in fact an antineutrino. The inverse reaction is  $\bar{\nu}_e + n \rightarrow p + e^-$  and there is the equivalent reaction with protons:  $\bar{\nu}_e + p \rightarrow n + e^+$ .

This last-mentioned is the process

detected experimentally by Reines and Cowan in the 1950s, leading to the universal acceptance of the neutrino (of no mass and no charge) as a *bona fide* particle. It was discovered later that there are other neutrinos too — most importantly here, the muon neutrino, generated in the decay of pions:  $\pi^+ \rightarrow \mu^+ + \nu_\mu$ . The corresponding inverse reaction creates muons:  $\nu_\mu + n \rightarrow p + \mu^-$  and it is this reaction that is the basis of the present excitement.

The search for muons produced by cosmic-ray neutrinos started in the early 1960s when the first deep-mine experiments commenced with what were then



Momentum spectra of electron-like and muon-like events observed in the Kamiokande experiment<sup>1</sup>. The histograms are the expectations for electrons and muons generated by the interactions of neutrinos, the neutrinos themselves having come from the interactions of cosmic rays in the atmosphere.

very big detectors. Our group working in the Kolar Gold Fields<sup>2</sup> and another in South Africa<sup>3</sup> detected the underground muons, some of which were moving upwards and thus must have been generated by neutrino scattering (at somewhat below the correct rate). A hypothetical particle of interest then was the 'intermediate vector boson' and a lower limit of 3 gigaelectron volts (GeV) was put on its mass from these results; as is well known, its discovery had to wait until the recent experiments at CERN led by Rubbia, and its mass is in fact nearly 90 GeV.

The search for the decay of the proton, expected on the basis of the grand unified theory, led to an upsurge in underground physics, and the Kamiokande experiment is one of two large-scale projects (the US IMB experiment is the other). It comprises a 3,000-ton water Cerenkov detector located in the Kamioka mine in Japan and has been in operation since July 1983. Particles produced in the water tank are detected by the Cerenkov radiation (electromagnetic 'shock waves') they produce and the pattern of the radiation leads to particle identification, albeit with some uncertainty. Hirata *et al.*<sup>1</sup> use Monte Carlo calculations to derive the expected number of electron-like and muon-like events (see figure). Clearly, if the electron-like events are indeed electrons, and similarly for the muons, then there is a deficit of muons whereas the number of electrons is just about correct.

Before interpreting this exciting result, one must ask some pertinent questions. Firstly, have the fluxes of cosmic-ray neutrinos been correctly calculated? Although it is true that at low energies there are problems, it is unlikely that there is much uncertainty above 500 MeV  $c^{-1}$ : the right-hand half of each plot in the figure should be all right (the recent flux calculations of Gaisser<sup>4</sup> give similar results to those we derived 20 years ago<sup>5</sup>; the uncertainty in flux is almost certainly no more than 20 per cent). The interaction cross-sections used to predict the detection rate are well known from accelerator experiments and these should present no problem. Thus, unless there are problems of identification of muons, with about half of them being lost, we must take the results seriously.

If correct, what is the interpretation? The most likely cause is the neutrino oscillation: the change in neutrino from one type to another *en route* from source to detector. Such behaviour, long sought for, has consequences not only for elementary-particle physics but also for cosmology, for which the associated requirement of a finite mass for the neutrino raises the possibility of gravitationally 'closing the Universe' through the summed mass of all the truly cosmic neutrinos left over from the Big Bang.

But if there is an oscillation between the muon and electron neutrinos, there should be an excess in the measured number of electrons produced to balance the loss of muons. This does not seem to be happening (see figure). It is possible that the lifetimes for oscillation are such that oscillations between the muon and tau-meson neutrinos (another version) are responsible.

Clearly, results from other experiments are needed. The early neutrino searches, with their suggestion of a deficit of muons, have been mentioned already. The preliminary analysis of the latest Kolar Gold Fields work is not encouraging, however: with more than one hundred muons collected to date the discrepancy has not been confirmed (V.S. Narasimham, personal communication). Hirata *et al.*<sup>1</sup> mention some evidence from the IMB and Frejus experiments indicating that the ratio of muon- to electron-neutrinos in their data is small. But the oscillation hypothesis will not be confirmed until a much more detailed analysis, now in progress, is made of the results from these and other experiments. □

- Hirata, K.S. *et al.* *Phys. Lett.* **B205**, 416–420 (1988).
- Acher, C.V. *et al.* *Phys. Lett.* **18**, 196–199 (1965).
- Reines, F. *et al.* *Phys. Rev. Lett.* **15**, 429–433 (1965).
- Gaisser, T.K. & Grillo, A.F. *Phys. Rev.* **D36** 2752–2756 (1987).
- Osborne, J.L., Said S.S. & Wolfendale, A.W. *Proc. phys. Soc.* **86**, 93–99 (1965).

A.W. Wolfendale is in the Department of Physics, University of Durham, Durham DH1 3LE, UK.

## Phylogeny

# Ancestry of sea mammals

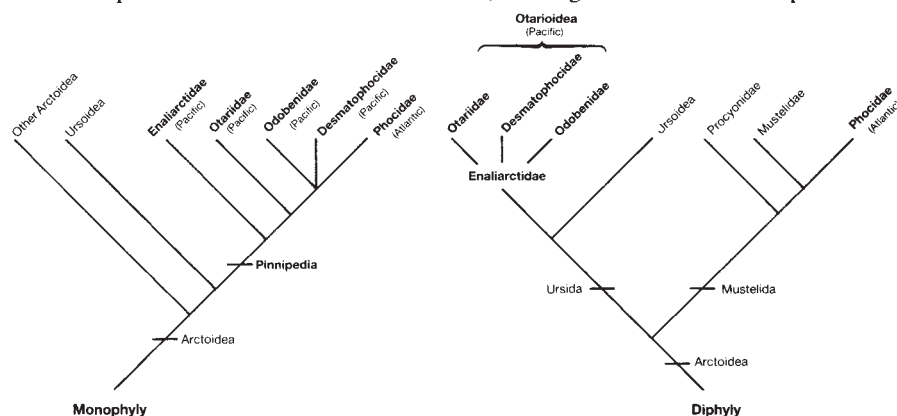
John J. Flynn

THE evolutionary relationships of marine mammals have been controversial and confusing since Aristotle<sup>1</sup> and Linnaeus<sup>2</sup> grouped whales and dolphins close to fish because of their aquatic specializations. In the history of terrestrial vertebrates, several groups have independently and secondarily become aquatic — including among the reptiles, the plesiosaurs, ichthyosaurs, sea snakes, sea turtles and mosasaurs; and among the mammals, the sea cows (Sirenia), dolphins and whales (Cetacea) and the Pinnipedia, which includes the seals (Phocidae), sea lions (Otariidae) and walrus (Odobenidae). This last group includes the members of the order Carnivora that are the most

specialized for an aquatic existence but there has been controversy about their origins. On page 427 of this issue, Wyss<sup>3</sup> provides compelling evidence that all the marine pinnipeds originated from a single terrestrial ancestral group.

Although the original classification of the seals, sea lions and walrus in a single group implied a single or monophyletic origin for these marine carnivores, the alternative hypothesis — that there was parallel evolution from two independent ancestral terrestrial groups — has recently become widely accepted among neontologists and palaeontologists (see below). This diphyletic scheme proposes that seals originated from weasel- and otter-like

The diphyletic origin of the pinnipeds was accepted primarily because of evidence from fossils, biogeography and divergence in the groups. First, fossils were identified that were proposed to represent intermediate stages of evolution between musteloids and seals on the one hand and ursoids and 'otarioids' on the other. Second, the geographical distribution led to suggestions of restricted origin and diversification of phocids in the Atlantic and 'otarioids' in the Pacific. Lastly, phocids were considered highly specialized and widely divergent, both anatomically and phylogenetically, from other pinnipeds. Recently, support for diphyly has been declining. The evidence that phocids are more closely related to musteloids than to other pinnipeds is weak<sup>6,7</sup>. Biochemical evidence also generally supports monophyly rather than diphyly<sup>8,9</sup>. Biogeographic evidence is equivocal, whether one proposes monophyly or diphyly. Several studies<sup>5,10</sup> indicate that the Pacific 'Otarioidae' are not a monophyletic group; some, such as the Enaliarctids, represent basal pinnipeds, whereas walrus and fossil desmatophocids are more closely related to phocids than to otariids. Therefore, although the oldest known phocids are



restricted to the Atlantic, their closest relatives are Pacific in distribution. There is not a simple Pacific/Atlantic dichotomy to support a diphyletic origin of 'otarioids' in the Pacific and phocids in the Atlantic. Further, the most conservative, basal phocid (*Monachus schauinslandi*) is found in the Pacific, north west of Hawaii, suggesting that even the idea of phocid origin in the Atlantic may be incorrect<sup>11,12</sup>. Features unique to a group tell us nothing about relationships of that group to any other; the specializations of seals would make them different from all other groups, including musteloids, and would not support or contradict either diphyly or monophyly. Further, recognition of the morphological features that show the close relationship of the walrus and desmatophocids to the seals greatly reduces the supposed morphological gap between phocids and other pinnipeds<sup>6</sup>. It is difficult to hypothesize a common ancestor of this closely related group, as required by diphyly, which would be more similar to musteloids than to otariids. Finally, perhaps the strongest evidence against diphyly is the support for pinniped monophyly provided by Wyss<sup>3</sup>.





Francisco Ortiz/Bruce Coleman

Not-so-distant relations? The Atlantic walrus *Odobenus r. rosmarus* (top) in the Canadian Arctic, and the northern fur seal *Callorhinus ursinus* on Pribilof Island in the Bering Sea.



forms (Musteloidea), whereas sea lions and walruses, the 'Otarioidea' came from bear-like forms (Ursoidea)<sup>3-5</sup>.

Diphyley requires that most of the similarities between pinnipeds are convergent, that is, they are independently derived but appear similar because of the intense selection pressure of their common aquatic environment. This 'functional or adaptive scenario' of convergence has, in fact, precluded use of almost any anatomical similarity between different pinnipeds as evidence of common ancestry and close evolutionary relationship. Wyss<sup>3</sup> makes a crucial point: anatomical similarities cannot be explained as convergent adaptations simply because of similar function or habitus. Convergence is relevant only in the context of the phylogenetic relationships of the organisms being compared — to show that convergent features evolved independently, we must already know that groups are distantly related (diphyletic) and that their common ancestor did not possess those features.

Pinnipeds locomote in very different ways, yet all pinnipeds share uniquely derived anatomical details of the flippers and skeleton, as well as the cranium and soft anatomy. These derived details are not found in other aquatic mammals or in any of the presumed closest relatives in the terrestrial Carnivora and so cannot arbitrarily be explained away as functional convergences. Few people doubt that adaptation to aquatic life results in evolution of crudely similar structures even in distantly related animals, for instance, flippers and fusiform bodies in ichthyosaurs, mosasaurs, pinnipeds, cetaceans and sea cows. Yet, because of the distant phylogenetic relationship between these groups and the distinctly different details

of their anatomy, it is clear that there have been a variety of independent solutions to an aquatic existence.

The extensive cranial and skeletal specializations shared by all pinnipeds and no other vertebrate, aquatic or terrestrial, can best be explained by common heritage rather than common function. Apparently pinnipeds entered the sea only once. Taxonomic diversity, biogeography, morphology, aquatic adaptations and evolutionary patterns in the Pinnipedia must now be explained within the context of a single origin of the group. □

John J. Flynn is in the Department of Geology, Field Museum of Natural History, Chicago, Illinois 60605, USA.

### Acetylcholine receptor

## Genes encoding nicotinic receptor subtypes on neurons

Darwin K. Berg and Stanley W. Halvorsen

NICOTINIC acetylcholine receptors (AChRs) in muscle and in the electric organ of eels are well characterized and provide the best model for understanding receptor function and regulation at synapses. But surprisingly little is known about their nature and role in neurons. Membrane components that bind nicotine with varying degrees of affinity have been identified in vertebrate and invertebrate nervous systems, but their relationship to functional receptors is, in most cases, obscure. At a recent meeting\* progress in identifying gene families encoding functional receptors was reviewed. The findings indicate that AChRs are widely distributed in the nervous system, that several receptor subtypes exist, and that more are likely to be found. It is now possible to examine the functional significance and properties of specific receptor subtypes in the nervous system.

Muscle and electric organ AChRs are pentameric with four types of subunit ( $\alpha_2\beta\gamma\delta$ ). Complementary DNA and genomic clones can distinguish at least three types of non-muscle receptor  $\alpha$ -genes and one type of non-muscle receptor  $\beta$ -gene that are expressed in rat (S. Heinemann, Salk Institute) and chick brain (M. Ballivet, University of Geneva). The neuronal genes are termed  $\alpha_2$ ,  $\alpha_3$ ,  $\alpha_4$  and  $\beta_2$ , respectively, with  $\alpha_1$  and  $\beta_1$  referring to the muscle receptor homologues. Expression studies with *Xenopus* oocytes demonstrate that functional receptors can be constructed from the combinations of  $\alpha_2/\beta_2$ ,

$\alpha_3/\beta_2$  and  $\alpha_4/\beta_2$ , and that their pharmacology is as expected for neuronal rather than muscle receptors. The fact that only two types of subunit are necessary to establish receptor function is consistent with reports that purified receptors from brain (J. Lindstrom, Salk Institute) and autonomic ganglia (S. W. H.) reveal only two types of subunit per receptor subtype. Either neuronal receptors differ in this respect from those in muscle and electric organ, or additional subunits remain to be discovered in neuronal receptors.

*In situ* hybridization demonstrates that the identified neuronal AChR genes are expressed in at least 70 per cent of about 300 distinct brain regions, suggesting that nicotinic transmission is widespread in the vertebrate central nervous system (Heinemann). The  $\beta_2$  gene product could contribute to several receptor subtypes because the distribution of  $\beta_2$  transcripts overlaps most of that for each of the identified neuronal  $\alpha$ -gene transcripts. Additional neuronal receptor genes that are also likely to exist as  $\beta_2$  transcripts occur in regions lacking known  $\alpha$ -gene transcripts, such as regions of the cerebellum, and vice versa, as in regions of the olfactory bulb. Genes tentatively identified as  $\alpha_5$  and  $\beta_3$  are currently being characterized in both rat and chick. Moreover, a gene cluster has been identified in chick involving an  $\alpha_5$ , an  $\alpha_3$  and a related non- $\alpha$ -gene (different from  $\beta_2$  and  $\beta_3$ ) in a contiguous arrangement with the latter two genes having the same orientation (Ballivet). The gene cluster poses interesting regulatory opportunities.

1. Aristotle *Historia Animalium* (trans. D. Thompson) 486-683 (Clarendon, Oxford, 1910).
2. Linnaeus, C. *Systema naturae, sive Regna tria Naturae systematice proposita per Classes, Ordines, Genera, & Species* (tolp. Lugduni Batavorum, 1735).
3. Wyss, A.R. *Nature* 334, 427-428 (1988).
4. McLaren, I.A. *Syst. Zool.* 9, 18-28 (1960).
5. Tedford, R.H. *Syst. Zool.* 25, 363-374 (1976).
6. Repenning, C.A., Ray, C.E. & Grigorescu, D. in *Historical Biogeography, Plate Tectonics, and the Changing Environment* (eds Gray, J. & Boucot, A.J.) 357-369 (Oregon State University Press, 1979).
7. Wyss, A.R. *Am. Mus. Novitates* 2871, 1-31 (1987).
8. Flynn, J.J., Neff, N.A. & Tedford, R.H. in *The Phylogeny and Classification of the Tetrapods, Volume 2: Mammals* (ed. Benton, M.J.) 73-116 (Clarendon, Oxford, 1988).
9. Arnason, U. & Widegren, B. *Molec. Biol. Evol.* 3, 356-365 (1986).
10. Miyamoto, M.M. & Goodman, M. *Syst. Zool.* 35, 230-240 (1986).
11. Berta, A. *Abstr. 7th Biennial Conf. Biol. Marine Mammals 5* (Soc. Marine Mammal., 1987).
12. Wyss, A.R. *Am. Mus. Novitates* (in the press).
13. Repenning, C.A. & Ray, C.E. *Proc. Biol. Soc. Washington* 89, 667-688 (1987).

\*NATO Advanced Research Workshop *Nicotinic Acetylcholine Receptors in the Nervous System*, Venice, Italy, 16-20 April 1988.

The availability of identified nicotinic AChR genes will allow a correlation with the host of binding components for nicotinic agonists of vertebrate brain (P. Clarke, University of British Columbia; A. Collins, University of Colorado; L. Costa, University of Washington; E. Perry, University of Newcastle). A nicotinic  $\alpha$ -bungarotoxin-binding component has been identified in brain (E. Barnard, MRC Cambridge; H. Betz, Molecular Biology Centre, Heidelberg; V. Chiappinelli, St Louis University; S. Wonnacott, University of Bath) and in neuronal cell lines (F. Clementi, University of Milan; G. Kemp, University of Alabama), but its function remains unknown. Purification of the component from chick brain reveals several subunits, and a partial amino-acid sequence reported for one of the subunits differs from the AChR genes for which deduced amino-acid sequences are available ( $\alpha 1$ – $\alpha 4$ ,  $\beta 1$ ,  $\beta 2$ ,  $\gamma$  and  $\delta$ ; Barnard). An  $\alpha$ -bungarotoxin-binding component in the human cell line TE671 seems to represent a muscle AChR (Lindstrom; R. Lucas, Barrow Neurological Institute).

Striking results have been obtained with neuronal receptors from insects. Locust receptors, affinity-purified with  $\alpha$ -bungarotoxin, have been reconstituted in planar lipid bilayers to obtain a nicotinic response (H. Breer, University of Stuttgart-Hohenheim). Monospecific antibodies against the purified receptor indicate that *in vivo* it is located in the ganglionic neuropil. Analysis of the purified receptor by SDS-PAGE reveals a single protein component of relative molecular mass 65,000 (65K), and reconstitution experiments with the 65K band in artificial bilayers produce a nicotinic response similar to that obtained with the purified receptor before denaturation. A nicotinic response is also obtained in expression studies with *Xenopus* oocytes injected with RNA from a locust complementary DNA clone (J. Marshall, MRC Cambridge), demonstrating that a single insect gene product can produce a functional receptor. Single-channel recordings show, however, that the predominant channel size of purified receptors after reconstitution is 75–80 pS, instead of the 40–50 pS channel observed for receptors on insect neurons *in situ* (Breer; D. Sattelle, MRC Cambridge). Either the reconstituted receptors differ in some respect from native receptors or receptor subpopulations with different properties exist on the neurons, for example, synaptic and extrasynaptic receptors.

New information on the function and regulation of neuronal receptors is now being obtained. Quantitation of receptor-site densities on vertebrate neurons shows that the receptors are concentrated at synapses but the packing is 3- and 25-fold less dense at synapses on rat superior

cervical ganglion neurons and chick ciliary ganglion neurons, respectively, than at the rat neuromuscular junction (R. Zigmond, Harvard Medical School). Other studies indicate that, in contrast to muscle receptors, the function of chick ciliary ganglion receptors can be augmented by a cyclic AMP-dependent process, and a cyclic AMP-dependent protein kinase can specifically phosphorylate the agonist-binding subunit of the receptor (D.K.B.). The functional relevance of the phosphorylation remains to be determined.

Several reports underscore the medical relevance of nicotinic receptors. Patients

suffering from Alzheimer's disease have reduced numbers of nicotine-binding sites in brain samples (E. Giacobini, Southern Illinois University). Positron emission tomography of human subjects indicates that smokers may be much more efficient at accumulating and retaining nicotine in their brains than non-smokers (A. Nordberg, University of Uppsala). The prospects are encouraging for elucidating nicotinic events in the nervous system at the cellular and molecular levels. □

*Darwin K. Berg and Stanley W. Halvorsen are in the Department of Biology, University of California, San Diego, La Jolla, California 92093, USA.*

## Ocean Drilling Program

# That sinking feeling

### Leg 120 shipboard scientific party\*

LEG 120 is the second of a two-leg drilling programme on the Kerguelen Plateau and completes a latitudinal transect in the Southern Ocean between Kerguelen Island (49° S) and Prydz Bay, Antarctica (67° S). The two Legs studied the nature and age of the Kerguelen Plateau basement, the origin and tectonic history of the Kerguelen Plateau, and the Late Cretaceous to Holocene oceanographic history (from about 90 million years ago to the present) of this region. Leg 119 drilled the northernmost (736 and 737) and southernmost (738 and 744–746) Kerguelen Plateau sites and Leg 120 drilled the central and southern plateau sites (747–751) (see figure). Leg 120 clearly established the basaltic nature of the Southern Kerguelen Plateau basement rocks and revealed the intermittent subsidence that occurred during the Cretaceous.

Located in the south-central Indian Ocean, the Kerguelen Plateau is one of the world's largest submarine plateaux. It is bounded to the north-east by the Australian–Antarctic Basin, to the south by the 3,500-m-deep Princess Elizabeth Trough, to the south-west by the African–Antarctic

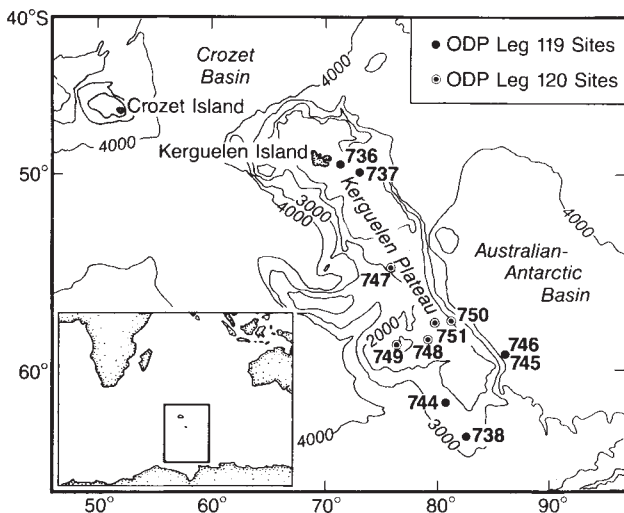
Basin, and to the north-west by the Crozet Basin. The plateau stretches approximately 2,300 km between 46° S and 64° S in a north-west–south-east direction towards the Antarctic continental margin and is between 200 and 600 km wide, standing 2–4 km above the adjacent ocean basins. It lies south of the present-day polar front (Antarctic Convergence) and beneath the main flow of the Antarctic Circumpolar Current.

The Kerguelen Plateau has been divided into two distinct domains<sup>1,2</sup>. The Northern Kerguelen Plateau, or Kerguelen–Heard Plateau, located between 46° S and 54° S, generally lies in water depths less than 1,000 m, and includes the plateau's only features above water: Kerguelen, Heard and McDonald islands. The Southern Kerguelen Plateau, located between 57° S and 64° S, is deeper, generally lying in water depths of 1,500–2,000 m, and has a subdued topography. The transition zone has complex bathymetry with a large east–west-trending spur, the Elan Bank, extending westwards from the main plateau over 600 km.

Except at site 748, where there is a silica-undersaturated alkaline lava flow (ocean island basalt-type), the basement basalts are silica-saturated transitional tholeiite (transitional mid-ocean ridge basalt). These seem to represent a mantle source of intermediate composition between depleted spreading ridge and enriched hotspot mantle reservoirs. The incompatible-element contents and ratios of basalts from the Kerguelen Plateau basement vary between sites, possibly indicating differences in degrees of partial melting and source heterogeneities. But these differences are small compared with those of basalts erupted at ocean ridges and intra-plate islands. All the basalts contain secondary minerals indicating hydrothermal alteration, from the

\* Co-Chief Scientists: Roland Schlich (Inst. de Physique du Globe, Strasbourg) and Sherwood W. Wise Jr (Geology Dept., Florida State Univ.). ODP Staff Scientist: Amanda A. Palmer (ODP, Texas A&M Univ.). Also: M.-P. Aubry (Univ. Claude Bernard), William A. Berggren (Woods Hole Oceanographic Inst.), Peter R. Bitschene (Mineralogisches Inst., Ruhr Univ. Bochum), Neal A. Blackburn (Britoil p.l.c.), James Breza (Florida State Univ.), Millard Coffin (Bureau of Mineral Resources, Canberra), David M. Harwood (Byrd Polar Research Center, Ohio State Univ.), Franz Heider (Univ. Toronto), Mary Anne Holmes (Univ. of Nebraska), William R. Howard (Brown Univ., Providence), Hiroo Inokuchi (Kobe Univ.), Kerry R. Kelts (EAWAG/ETH, Zurich), David Lazarus (Woods Hole), Andreas Mackensen (Alfred Wegener Inst. for Polar Research, Bremerhaven), Toshiaki Maruyama (Tohoku Univ.), Marc Munsch (Inst. de Physique du Globe, Strasbourg), Elizabeth Praton (Lamont-Doherty Geological Obs.), Patrick G. Quilty (Antarctic Div., Tasmania), Frank Rack (ODP, Texas A&M Univ.), Vincent J.M. Salters (Massachusetts Institute of Technology), James H. Sevigny (Univ. Calgary), Michael Storey (Univ. Leicester), Atsushi Takemura (Kyoto Univ.), David Watkins (Univ. Nebraska), Hubert Whitechurch (Inst. de Geologie, Strasbourg) and James C. Zachos (Univ. Rhode Island).





Bathymetric chart of the Kerguelen Plateau showing Leg 119 and Leg 120 drilling locations.

low-temperature (sites 747 and 750) to the high-temperature (749) zones of the zeolite facies.

The evolution of the Southern Kerguelen Plateau can be summarized as follows. First, basalts were erupted before the Cenomanian (98 Myr) above or just below the sea surface and were emplaced on a near-horizontal surface. The basalt flows are relatively thick and massive in the eastern part of the Ragatt Basin (site 750), fresh to highly altered on the Banzare Bank (site 749) and strongly weathered and interlayered with siltstones and claystones produced by weathering in the western part of the Ragatt Basin (site 748). The early sedimentation denotes fluvial conditions and consists at site 750 of a broad range of water-laid terrigenous claystones and siltstones, with some sandy or conglomeratic intervals. Similar sediments were found at site 748. Wood fragments denote the development of soils and vegetation on some of these flows.

During the Cenomanian, Turonian and, probably, Santonian (between 98 and 83 Myr), sedimentation evolved in the eastern part of the Ragatt Basin (site 750) to open marine deposits of chalk with dark clay interlayers. The eastern margin of the plateau slowly subsided to about 50 m below sea level (m.b.s.l.). To the west (site 748) the plateau remained at about sea level, with sedimentation of glauconitic sand-, silt- and claystones, no or rare silicified bioclastic debris and some pyritized wood fragments.

During the Campanian and late Maastrichtian (83–65 Myr), the eastern margin of the Southern Kerguelen Plateau subsided rapidly to about 2,000 m.b.s.l. and open marine sedimentation continued

at a rate of 10–20 m  $\text{Myr}^{-1}$  with nanno-fossil chalk, chert and intermittently silicified limestone. To the west the plateau subsided only slowly and remained shallow (50–200 m.b.s.l.). The sediments (deposition rate  $\sim 46 \text{ m Myr}^{-1}$ ) consist of glauconitic sand-, silt- and claystones, having evolved progressively to intermittently silicified rudstones, grainstones and wackestones.

At about 75 Myr a major tectonic episode affected the eastern margin of the

Southern Kerguelen Plateau (site 750). To the west this tectonic event was recorded at about 68 Myr (site 748). Normal listric faults developed to the east (site 750) and are possibly related to the break between the Southern Kerguelen Plateau and Broken Ridge–Diamantina fracture zone. To the west (site 748) this event corresponds to the emplacement of the 77° E Graben and to a large north-west–south-east uplift that abuts the southern end of the 77° E Graben. At this time the western margin of the Ragatt Basin (site 748) subsided rapidly to 1,000 m.b.s.l., at a rate of about 150 m  $\text{Myr}^{-1}$ .

From the late Maastrichtian to the middle Eocene (about 45 Myr), sedimentation (mainly nannofossil chalk and ooze with some chert) was continuous all over the plateau at a rate of about 18 m  $\text{Myr}^{-1}$  to the west (site 748) and of 5–30 m  $\text{Myr}^{-1}$  to the east. The plateau subsided only slowly.

A hiatus of at least 2 Myr occurred during the middle Eocene (sites 748 and 750). At site 747, in the transition zone between the Northern and Southern Kerguelen Plateau, this hiatus extends over 15 Myr and is accompanied by a subsidence of about 500 m. This event can be related to the separation by seafloor spreading<sup>1</sup> of the Northern Kerguelen Plateau and Broken Ridge dated at 43–42 Myr. From the middle Eocene to the Pliocene (5 Myr), sedimentation continued at up to 15–20 m  $\text{Myr}^{-1}$  all over the plateau and resulted mainly in nannofossil ooze. During the Pliocene and the Pleistocene (until 100,000 years ago) a low sedimentation rate of 3 m  $\text{Myr}^{-1}$  is recorded in the central part of the Ragatt Basin (site 751) and is related to erosion that affected the whole Southern Kerguelen Plateau. □

## Daedalus

### Beaming from on high

THE geostationary orbit is inconveniently high for relay satellites. A TV signal of a mere few watts, beamed from 36,000 km up, can be picked up only by an expensive dish aerial. The obvious solution, that of spinning the Earth up to an 86-minute day, thus bringing the geostationary orbit down to a convenient 70 km altitude is a bit extreme, even for Daedalus. He has a simpler way of suspending a communications relay at that height.

He has been inspired by the Crookes' radiometer, that little windmill in a glass flask which spins in sunlight. The radiation warms the blackened side of each vane, which pushes the adjacent gas away from that side by thermal transpiration effects. The resulting reaction propels the vane away from the radiation. The effect is greatest at a pressure of around 5 pascals or so, just about the pressure at 70 km altitude. A big, light, convex reflector might easily be levitated at that height by the heating effect of microwaves beamed up to it from a powerful transmitter on the ground. The energy not absorbed in thermal levitation would be reflected back to Earth in a widely diverging beam. Careful shaping of the reflector could define the desired 'reception footprint' exactly. It could extend to an ultimate horizon nearly 1,000 km away.

One complication is that the levitated mirror would not be stable. If it drifted sideways from the beam centre, the thermal-transpiration forces would push it further sideways still. The lifting beam would have to juggle with it constantly, holding it in the sky like a plate balanced on an invisible stick.

The electronics required for this trick would be trivial. All the functions of a relay satellite would still be achieved at a fraction of the cost. Strong TV signals, data and telephone traffic could all be uplinked by the main levitating transmitter and relayed by reflection over most of a continent. Other transmitters within the footprint could also bounce their signals off the big sky mirror, provided they were not powerful enough to push it seriously off station. But how to get it up there in the first place?

Cunningly, Daedalus will fashion his reflector as a light, aluminized, plastic-film balloon, able to rise the first 40 km under its own helium. The levitating microwave beam will then be aimed at it, lifting it as a 'hot-helium' balloon until it is high enough for thermal transpiration forces to take over. It will soon rise to its stabilized position, reflecting the wonderful nonsense of modern TV entertainment, bureaucratic data traffic and compulsive telephone chatter strongly and bountifully over its whole vast reflection area. David Jones

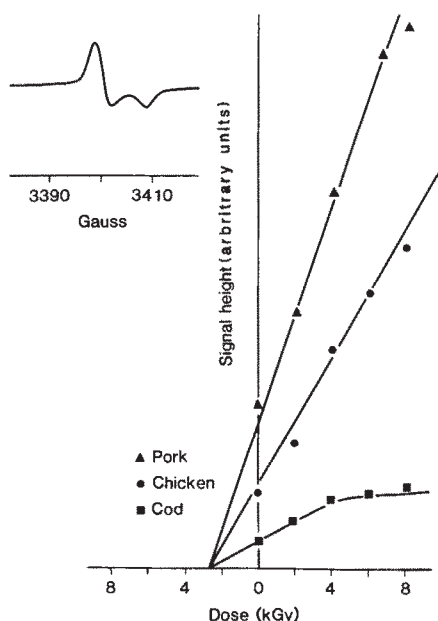
## Erratum

The onset of glaciation in Antarctica occurred during the Eocene–Oligocene, not the Miocene as stated at the end of the first paragraph of the Leg 119 report (*Nature* 333, 303–304; 1988).

- Schlich, R. *Soc. geol. France, Mem. Hors-Serie*, 6, 1–103 (1975).
- Houtz, R.E. *et al. Mar. Geol.* 25, 95–130 (1977).
- Munschy, M. & Schlich, R. *Mar. Geol.* 76, 131–152 (1987).

## ESR detection of irradiated food

SIR—A wide range of methods for the detection of irradiated food<sup>1</sup> is currently under investigation, as it is unlikely that one method will be applicable to all food-stuffs. At present, only thermoluminescence<sup>2</sup> and electron spin resonance (ESR)<sup>3</sup> are well developed. When applied to irradiated seeds or spices, ESR has no advantage over thermoluminescence, as the signal is generally non-specific and decays rapidly, the rate of decay being dependent on moisture content, thus preventing estimation of dose. On the other hand, as we describe below, samples containing bone or other calcified tissue, such as the shells of molluscs or crustacea,



Dose-response curves for samples of the three species. The initial dose at 2 kGy is shown as the origin with subsequent re-irradiation doses of 2–8 kGy. Extrapolation back to the axis reveals the experimental estimate of the original dose. Inset: typical e.s.r. spectrum of bone.

show an ESR signal that is both stable and characteristic of irradiation.

We irradiated samples of pork, chicken and cod with a dose of 2 kGy, using a 10 MeV electron linear accelerator. The bones were cleaned, powdered and dried and their ESR spectra recorded using a Varian E-9 X-band spectrometer. The spectra vary somewhat between species but all show the characteristic signal of an axially symmetrical radical (see figure), attributed to  $\text{CO}_2^-$  trapped in a hydroxy-apatite matrix<sup>4</sup>. Unirradiated bones showed no detectable signal under the same conditions and thus ESR clearly indicates whether the samples have been irradiated. The height of the signal is greatest in the highly calcified pork, intermediate in the chicken and lowest in the poorly calcified fish bone. Smaller variations in signal height have been observed between samples of the same species.

To estimate the dose received by a sample of bone, it is necessary to re-irradiate to a known dose. The figure shows the signal heights obtained for each sample after re-irradiation with doses of 2, 4, 6 and 8 kGy. Extrapolation to zero signal would indicate that the pork and chicken samples had originally received a dose of  $2.6 \pm 1$  kGy, in agreement with the actual dose. The sample of fish bone shows a non-linear dose-response curve at higher doses, but extrapolation from the lower doses again indicates an original dose of about 2.6 kGy.

For this test to give reliable quantitative results in practice, it is important that there should be no significant decay of the radiation-induced signal under normal conditions of processing and storage. Our work on chicken has shown that there is no significant decay over three weeks at 4°C or several months at -21°C, while the signal from the dried powders is stable indefinitely. Moreover, cooking after irradiation produces no significant loss of signal. The lower limit of detection is 50 Gy. This limit will be higher in fish bone but still below the minimum dose likely to be employed commercially.

These results demonstrate that ESR is a valid method for the detection and quantification of irradiation in a broad class of foodstuffs containing bone or other calcified tissue. It might also be used for other dry foodstuffs such as seeds and spices, its accuracy being comparable to that of thermoluminescence.

Fear has been expressed at the possible harmful effects of radiation-induced free radicals, but radicals *per se* are not harmful and many natural biological processes involve radical formation. Human life itself would be impossible without that ubiquitous radical molecular oxygen. The presence of long-lived radicals in irradiated foodstuffs<sup>5</sup>, provides no cause for concern, because they appear either to be indistinguishable from naturally occurring radicals or are trapped in inedible parts of the food.

N.J.F. DODD

J.S. LEA

A.J. SWALLOW

Paterson Institute for Cancer Research,  
Christie Hospital and Holt Radium  
Institute,  
Manchester M20 9BX, UK

1. Delincée, H., Ehlermann, D.A.E. & Bögl, K.W. in *Health Impact, Identification and Dosimetry of Irradiated Food* (eds Bögl, K.W., Regulla, D.F. & Suess, M.J.) 58–127 (Institut für Strahlenhygiene des Bundesgesundheitsamtes, Munich, 1988).
2. Moriarty, T.F., Oduko, J.M. & Spyron, N.M. *Nature* **332**, 22 (1988).
3. Dodd, N.J.F., Swallow, A.J. & Ley, F.J. *Radiat. Phys. Chem.* **26**, 451–453 (1985).
4. Geoffroy, M. & Tochon-Danguy, H.J. *Calcif. Tissue Int.* **34**, 99–102 (1982).
5. Webb, T. & Lang, T. in *Food Irradiation the Facts* (Thomson, Wellingborough, 1987).

## Polymerase chain reaction reveals cloning artefacts

SIR—The recently developed polymerase chain reaction (PCR) makes possible the *in vitro* amplification of specific DNA segments bounded by an oligonucleotide primer on each strand<sup>1</sup>. During successive cycles of denaturation and extension of the primers by a heat-resistant DNA polymerase of *Thermus aquaticus* (*Taq* polymerase)<sup>2</sup>, large amounts of amplification product can be generated from a few copies of template DNA. This technique holds great promise for forensic science and molecular archaeology as well as population biology.

Because *Taq* polymerase may have an overall error frequency of 0.25% in typical amplification reactions<sup>2</sup>, there is concern that sequences determined from amplified products may contain artefacts. This is a particular worry in the case of ancient DNA preparations, which contain a wide variety of chemical modifications<sup>3</sup>. To address this issue directly, we designed primers for the amplification of two regions of the mitochondrial genome from the quagga, an extinct member of the horse family<sup>4</sup>. The DNA sequences are located in the genes for cytochrome oxidase I and NADH dehydrogenase I, respectively, and have been shown by conventional cloning to be identical to those of the plains zebra, the nearest living relative of the quagga, except for two substitutions, one in each of the two clones<sup>4,5</sup>. Because both these substitutions would cause amino-acid replacements at positions which are conserved among all other vertebrates tested, it has been speculated that they represent cloning artefacts caused by post-mortem damage to the old DNA<sup>5</sup>.

PCR amplifications from DNA extracts of the same tissue specimen used for earlier cloning yielded products of the expected size. Single-stranded DNA of these amplification products was generated by the unbalanced priming method<sup>6</sup> and sequences were determined by the dideoxynucleotide method. The quagga sequences generated by PCR were identical to the cloned quagga sequences in all positions compared except the two discussed above. At these sites, the PCR products were identical to the orthologous plains zebra sequences (see figure). This confirms the previous assumption that these positions in the cloned sequences represent cloning artefacts.

Although nothing is known about the tolerance of *Taq* polymerase to various types of damage in the template DNA, it can be speculated that the enzyme, like other DNA polymerases, will be slowed down at damage such as baseless sites<sup>7</sup>. Furthermore, many types of damage such as inter- and intramolecular cross-links,



Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

# Fluent mechanics

Paul Fletcher and Philip T. Smith

**Linguistics: The Cambridge Survey, Vols 1-4.** Edited by Frederick J. Newmeyer. Cambridge University Press: 1988. Vol. 1 Linguistic Theory: Foundations. Pp.500. £27.50, \$49.50. Vol. 2 Linguistic Theory: Extensions and Implications. Pp.320. £27.50, \$39.50. Vol. 3 Language: Psychological and Biological Aspects. Pp.350. £27.50, \$39.50. Vol. 4 Language: The Socio-Cultural Context. Pp.294. £27.50, \$39.50. In Britain each volume £30 from 1 October.

MANY sciences attempt to understand phenomena that are beyond normal human capability: heavier-than-air flight, communication over large distances, microscopic vision and so on. Linguistics is the study of what most human beings are very good at — learning and using language — yet the puny mechanisms suggested by psychology and artificial intelligence seem quite incapable of emulating these skills. Such emulation must be the primary goal of linguistics; mere description of linguistic phenomena, though difficult enough in itself, inevitably falls short of true understanding, just as centuries of dressing up in feathers and jumping off cliffs failed to shed much light on the mechanisms of flight.

More than anyone else, Noam Chomsky has been responsible for promoting the view that human linguistic abilities form the core of linguistics, that it is the internalized grammar in the human mind that we should be studying, not its external manifestations in physical utterances and texts. This mentalism has not been without its detractors, both from within linguistics and from psychology.

Thus, there are hard-headed empiricists who believe it is dangerous to go beyond the immediate physical data: their position has been strengthened in recent years by the growth of computer data bases, where it is possible to search rapidly through millions of words to find instances of particular linguistic structures, so avoiding some of the less objective methods (appeals to intuition, for example) that linguists are often obliged to use. There are other linguists who perceive linguistics as a self-contained enterprise independent of other human sciences such as psychology, and there are psychologists who see language as just one manifestation, among many, of more general human cognitive abilities. For both the latter two groups, the failure in the 1970s of Chomsky's so-called transformational grammar to provide insights for psychological models of language learning and processing was evidence that linguists and psychologists should pursue their enterprises separately.

Recent developments involving Chomsky's work on government and binding, and psychologists' interest in

the potentials of modular systems, have revitalized the notion of linguistics as a psychological theory and are admirably reviewed in *The Cambridge Survey*. To give the flavour of current chomskyan linguistics, we dwell in a little detail on one linguistic phenomenon in English, the structure of interrogatives.

We begin with the observation that the sentence *What did you put in the garage?* requires, for its proper interpretation, that *What* is construed as the object of *put*. The structure of this sentence is then characterized as having resulted, by a process of *wh*-movement, from an underlying structure which we may represent as *You put wh in the garage*. The claim that the grammatical structure of a sentence is a complex entity requiring non-observable constructs is a commonplace of the chomskyan approach. Notice that the *wh*-word can be moved to the front of a sentence in a more complex structure than the one above — *What did Mary say that Bill put in the garage?* Here the verb *put*, to which *what* has to be related in order to understand the sentence as a whole, is 'embedded in' a further sentence. Is it always possible to 'extract' *wh*-elements from smaller sentences within complex structures? Obviously not, because we cannot say *Who were you embarrassed because you talked to?* (cf. *You were embarrassed because you talked to Mary*).

Observations such as this are central to the grammatical enterprise, and chomskyan grammarians have elaborated a set of principles which constrain the operation of grammatical rules like *wh*-movement under specific conditions. By using this interaction between grammatical rules and constraining principles (many of which can be seen to operate across languages) linguists have had a considerable success in accounting for a range of complex grammatical phenomena. The cross-linguistic applicability of theoretical linguistic constructs has, in the 1980s, restored some plausibility to the claim that the principles that the grammarian uses to account for the sentences of a language are exactly those that children use in the learning of their native tongue.

These ideas are prominent in the volumes under review. Most of Vol. I of

the *Survey* is concerned with detailed accounts of current thinking within core linguistics, and is organized by linguistic levels — phonetics, phonology, grammar and so on. There is however, via chapters on the interface between syntax and semantics, and between syntax and phonology, exploration of the ways in which adjacent levels are interdependent. The conclusion (in Vol. I on purely linguistic grounds) is that such levels are generally independent of one another in their operation, that is to say they are modular, and this theme of modularity is echoed in the more psychologically orientated evidence presented in subsequent volumes. Although no concessions whatever are made in Vol. I to non-specialist readers, the overall standard of articles is high. Those by Anderson on morphology, Horn on pragmatics and Lightfoot on syntactic change are particularly well organized.

Volumes II and III are concerned with the application of linguistic principles to psychological issues, and from a core of work on language learning, production and comprehension consider such diverse topics as creole languages and animal communication. Some of the contributors are treading well-worn paths that they themselves created, but there are masterly reviews by Garrett on language production and by the Gleitmans and their co-workers on first-language learning.

Volume IV extends the scope of the *Survey* from language as a mental or biological entity to the study of language in its sociocultural context. This role is very broadly conceived, and the 14 chapters cover a disparate set of topics. Those on language and race, language and gender, language death and language planning, for example, seem to fall squarely within the stated theme of the volume. Other topics, such as discourse organization and conversational analysis, sit here uneasily, and underline the arbitrary restriction (implicit in the organization of the *Survey* as a whole) that limits the consideration of language as a biological entity to phenomena at the sentence level or lower. The suspicion of thematic incoherence in this volume is strengthened by the inclusion of a chapter on bilingualism. Why, the reader may ask, is bilingualism any more or less of a sociocultural matter than monolingualism? Thematic problems apart, it is the chapters in this volume which will most engage the general reader.

As a survey of a wide range of research and practice in modern linguistics and related disciplines, unified by a particular theoretical perspective, the first three volumes have much to commend them. They convey the complexity of modern linguistic theory, and the intellectual excitement of its extension into areas such as language acquisition and language



processing that might have been thought to be more properly the province of psychologists.

On the success of the enterprise, however, the jury is still out. As Weinberg comments in Vol. I, "... a theory that characterizes what we know about our language in such a way that this knowledge could never efficiently be put to use would not seem to be a very promising psychological candidate". This cautionary note, written in the context of whether chomskyan grammars could be efficiently

parsed (that is, processed) by a machine, might be applied throughout these volumes. It is one thing to enunciate general constraints on language learning, it is quite another to incorporate these constraints into a realistic psychological model. But as a way of understanding complex phenomena, it still beats dressing up in feathers and jumping off a cliff. □

*Paul Fletcher is Reader in Linguistic Science and Philip T. Smith Reader in Psychology at the University of Reading, Whiteknights, Reading RG6 2AH, UK.*

## On fibreglass and fishes

*R. McNeill Alexander*

**Axis and Circumference: The Cylindrical Shape of Plants and Animals.** By Stephen A. Wainwright. *Harvard University Press: 1988. Pp. 132. \$22.95, £18.50.*

**The Science of Structures and Materials.** By J.E. Gordon. *W.H. Freeman: 1988. Pp. 217. \$32.95, £14.95.*

PROFESSOR Wainwright is a biologist who thinks like an engineer and Professor Gordon is an engineer with an interest in biology. Wainwright studies the design of animals and plants, thinking of them as structures built of appropriate materials. Gordon studies the design of structural materials, both man-made and natural. Both have written books that will make their ideas accessible to anyone who studied science at school. Wainwright's book should also be required reading for students and research workers in animal and plant morphology.

Wainwright has been remarkably influential as a teacher as well as in research. Most North American research workers in comparative biomechanics seem at some stage to have been his students, and to have been infected by his enthusiasm and mesmerized by his luxuriant bow ties. The theme of his new, short book is that the cylinder is the basic shape of multicellular life. Worms and eels are cylinders with pointed ends; rose bushes and monkeys are branched structures built of several cylinders each. A few basic principles associated with cylindrical shape illuminate most animal and plant structures.

To make a cylinder strong but flexible you should put its skeletal material along the axis, as in the roots of plants and the backs of vertebrate animals. To make it resist bending you should put the skeletal material at the circumference, as in bamboo stems and crab legs. To enable it to bend without kinking or twisting you should wrap it in a criss-cross array of helical fibres, like the fibres in the cuticle

of some nematode worms and under the blubber of whales. Joints and suitably arranged muscles make mechanisms built of rigid cylinders, such as human arms, capable of a wide range of movements. Muscles in flexible cylinders such as octopus arms can also produce a wide range of movements, free of the constraint of localized joints.

Occasionally, Wainwright's stimulating book seems just a little *too* easy to read. Not everyone who reads that "bending and twisting are complex and thus potentially dangerous deformations" will stop to wonder what, if anything, that means. A convincing-looking graph shows that a long, thin, cylindrical animal can make itself much longer by contracting its diameter slightly, but Wainwright fails to point out that a given percentage reduction in diameter always gives the same percentage increase in length, however stout or slender the cylinder.

Gordon took his first degree in naval architecture, and during the Second World War worked on materials for aircraft construction. Subsequently he

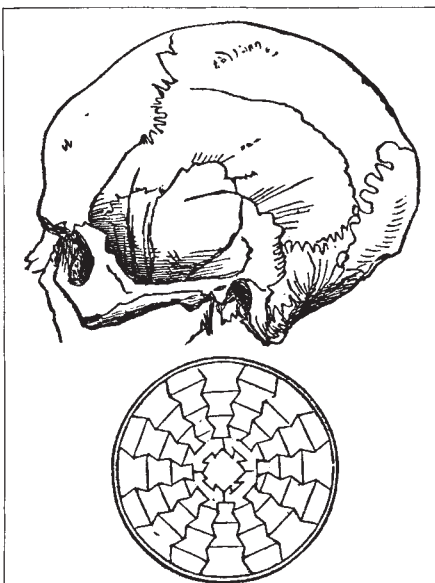
became a professor at Reading University, where he associated with biologists as well as with engineers. He played a distinguished part in the development of fracture mechanics but is also the author of the funniest and most enlightening popular books on mechanics, *The New Science of Strong Materials, or Why You Don't Fall Through the Floor* (1968) and *Structures* (1978). The main theme in this new book, as in the older ones, is that we all need to know about fracture mechanics, whether we be engineers, biologists, orthopaedic surgeons or even military historians.

The ancient Greeks may have won the Battle of Marathon because their annealed bronze armour was soft enough to be dented by the Persian arrows: it absorbed the arrows' kinetic energy and stopped them without cracking. Similarly, antler is more flexible than ordinary bone (it contains a smaller proportion of mineral salts) and so is better able to absorb the impacts of fighting stags. The same property made it excellent for making prehistoric pickaxes.

Antler, wood, fibreglass and the carbon-epoxy materials used in aircraft construction are composites, built of strong fibres embedded in a relatively weak matrix. The strength of the material is paradoxically dependent on the weakness of the matrix, which fails in front of the tip of a crack and dissipates the stress concentration that would enable the crack to spread. Roofs, bridges, aeroplanes and many other structures are best made as light as is safe. The weight of a structure can often be reduced by designing it so that as many as possible of its members are in tension, because compression



*Putting on the brakes—for a bird, large changes in geometry are needed to produce the forces required in manoeuvres such as landing: wings arched, legs outspread and tail bent forward and splayed to create drag while maintaining lift at low speed. Aircraft command much higher power/weight ratios and generally do not need such techniques, but aircraft like Concorde or the swing-wing fighter do display considerable changes in form. The picture is taken from a further book on structures, *Invention and Evolution: Design in Nature and Engineering* by M. J. French, recently published by Cambridge University Press, price hbk £35, \$59.50; pbk £12.50, \$19.95.*



Natural principle — dovetailing in the human skull and in the stonework of the Eddystone lighthouse (seen in section). The illustration first appeared in J.G. Wood's *Nature's Teachings: Human Invention Anticipated by Nature*, published in 1885.

members have to be made thick enough to prevent buckling. The lightest pillars can be made of materials of high Young's modulus and low density. Aluminium is better than steel, and magnesium would be better still were it not for the danger of fire, but wood and carbon fibre composites are better than any of the metals.

Discussions of these and similar topics fill Gordon's book, just as they filled his earlier ones. The message is the same but the presentation is glossier, with plenty of the apt colour photographs that we expect to see in the *Scientific American Library*. However, I found myself wondering from time to time whether Gordon's eccentric style had been curbed just a little by a well-meaning editor — *The New Science of Strong Materials* was more fun.

There is quite a lot of common ground between Wainwright's and Gordon's books. Wainwright shows us the geodetic fibres in the cuticles of worms and Gordon shows us the geodetic framework of the Wellington bomber. Both authors explain how the J-shaped stress-strain curve of many biological materials helps to prevent aneurisms, swellings such as occur, for example, in the early stages of inflating a sausage-shaped balloon. Both explain the Cook-Gordon mechanism of crack stopping, though Gordon does so in more detail (which is not surprising: he discovered it). Above all, neither author is conventional and both take us from basic principles to unexpected insights. To Gordon, the Parthenon is appallingly bad engineering, and to Wainwright a worm is a sock full of meat. □

R. McNeill Alexander is a Professor in the Department of Pure and Applied Biology, University of Leeds, Leeds LS2 9JT, UK.

## Events of no little matter

J.M. Charap

**The Great Design: Particles, Fields, and Creation.** By Robert K. Adair. Oxford University Press: 1988. Pp.376. £19.50, \$24.95.

SOMEWHERE between the paranormal and psychology, most bookstores will have a shelf or two of books on physics. Most of them are likely to be pot-boilers on black holes or particle physics, and few will attract the serious attention of professional scientists — which is not to say that they do not merit such attention. I have often wondered about the people who read these books. Presumably they will include the layman with the "attentive mind" to whom Adair addresses himself; what, then, will such a layman make of a book which is overtly pedagogic, which has a scope not very different from an American university non-calculus course in 'modern physics', and even slips some of the calculus into the end-of-chapter notes?

Well, I suppose many will be put off by the equations, and will not find the diagrams with their matchstick figures of Olivia and Oliver, the two independent observers, sufficiently enchanting to persevere with the text. This is a pity, because there is a lot to be said for this treatment, which is basically a straightforward, unfussy and well-ordered account of the core concepts of relativistic particle physics.

Most particle physicists would place conservation laws and invariance in this core. If pressed, most others would agree. And Adair puts these concepts right up front. He is very good in his treatment of special relativity, and even gets in a chapter on general relativity before introducing electromagnetism. This unusual order of presentation works, because he avoids the more conventional chronological approach in favour of a conceptually more logical sequence of ideas. The early emphasis on invariance and conservation laws, convincing in its own right, gives more point to the concluding chapters on unification and the links between cosmology and particle physics.

Although nearly half of the book is addressed to the concepts of particle physics as such, which derive so essentially from the revolutionary consequence of the marriage between relativity and quantum mechanics (namely, that particles can be created and destroyed), that central idea is not brought out as clearly as I think it should and could have been. Quantum mechanics is described as essentially a one-particle affair, and although later there will be a treatment of such detailed

topics as the modification of the universal Fermi interaction introduced by Cabibbo, or of the role of the Higgs field in the spontaneous breaking of symmetry in the electroweak theory, or of the renormalizability of gauge theories (!), one reads "in general, source (or matter) particles are conserved, and field (or force) particles are not". That is at best misleading, if not downright false.

It would not be possible to deduce from the text that Adair is Associate Director of the Brookhaven National Laboratory, for many years one of the centres of experimental high-energy and nuclear physics, nor that he himself is a distinguished experimentalist and phenomenologist. Little attempt is made to give more than a cursory description of the experiments from which the conceptual framework of particle physics derives; indeed, thought experiments are often given more prominence than those in the laboratory. This is a restrained and distanced account, one in which it is difficult to discern any sense of the joy and delight the author must find in his subject; his diffidence and good manners get in the way. For those who already have the motivation and the determination to follow a line of reasoning to its conclusion, Adair is on many topics an excellent guide and mentor, with a good teacher's ability to anticipate the difficulties the student will have.

I suppose, then, that the layman with the "attentive mind" will indeed read this book. If so, he will derive from it what the author had intended, namely an unadorned exposition of the basic concepts of particles and fields and of space and time, as they are understood and used by the professional physicist. That is no mean achievement. But I regret that the account is so sanitized, so lacking in passion and personality, in those humanizing aspects of the pursuit of science which might better have seized the mind and attention of the reader. After all the independent observers in the real world of physics are not matchstick Olivias and Olivers, but real men and women. □

J. M. Charap is a Professor in the Department of Physics and Pro-Principal of Queen Mary College, University of London, Mile End Road, London E1 4NS, UK.

### New in paperback

- *In Pursuit of the Past: Decoding the Archaeological Record* by Lewis R. Binford. Publisher is Thames and Hudson, price is £9.95, \$17. For review see *Nature* 303, 449 (1983).
- *The New Politics of Science* by David Dickson, with a new preface. Publisher is University of Chicago Press, price is \$14.95 (to be published in Britain later this year; originally published by Pantheon). For review see *Nature* 309, 813 (1984).
- *The History of Scurvy and Vitamin C* by Kenneth J. Carpenter. Publisher is Cambridge University Press, price is £9.95, \$12.95. For review see *Nature* 324, 177 (1986).



# Ringing out the old

Melvin G.R. Cannell

**Tree Rings: Basics and Applications of Dendrochronology.** By Fritz Hans Schweingruber. Reidel: 1988. Pp.279. Dfl 295, £84, \$149.

*Tree Rings* begins with an apt quotation: "Barriers erected between the different branches of knowledge are at the root of many of our problems". It was Dr Schweingruber's aim in writing the book to break down any barriers between dendrochronologists and people working in other disciplines.

His approach has been to produce a general survey of all the many facets of the subject, in a coffee-table style with numerous, self-explanatory illustrations. The illustrations are a striking feature, making it possible to delve into the book and quickly grasp the main ideas and information. No aspect is treated in depth, and readers may be irritated by oversimplification and the somewhat personal, subjective viewpoint. But, in my view, Schweingruber has achieved his aim.

Tree-ring research has advanced con-

siderably in recent decades; for example, we now have a good 7,000-year chronology for Europe. Such long-term biological records are rare and need to be fully exploited, verified and extended. This book shows that you do not have to be a plant physiologist, wood technologist or statistician in order to become involved, or at least to benefit from the work of dendrochronologists.

The book is divided into five sections. The first deals with the "Origin of the Materials" — that is, the trees and historic timbers from which samples are taken — and includes a simple discussion of site factors and sampling procedures. The second section, "Analysis of the Materials", covers coring techniques, methods of measurement and analysis, while the third describes some of the main findings, that

is, the sources of variation in ring widths and wood density. Next comes an intriguing survey of the ways in which dendrochronology has been applied to solve problems in, for instance, history, climatology, geomorphology, entomology, forestry and the environmental sciences. Finally, Schweingruber provides a personal review of the history of dendrochronology, complete with photographs of the luminaries.

The book is produced to a high standard, on high-quality paper, with clear figures and photographs. The text is of a comparable calibre, with many crisp sentences, and rarely betrays the translation from German. □

Melvin G.R. Cannell is Head of Station at the Institute of Terrestrial Ecology, Bush Estate, Penicuik, Midlothian EH26 8LA, UK.

## Small is beautiful

P.C. Turner

**Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.** Edited by Max L. Birnstiel. Springer-Verlag: 1988. Pp.216. DM 198, £68, \$120.

TO ANY molecular biologist who has the time to reflect, it will no doubt seem surprising that only a decade ago we were (blissfully?) unaware of the existence of the intervening sequences that are present in many of the genes of eukaryotic cells. What seems all the more amazing is how far we have come in that short time in understanding the processes of intron removal and of RNA processing in general. As a testament to how effectively the molecular biological approach has been applied to explore the catalysts involved in these phenomena, Max Birnstiel has put together an excellent collection of complementary reviews from several of the leading workers in the field. These comprehensive, readable articles will provide research students, teachers and specialists with a wealth of up-to-date information on this fast moving field.

The classification, occurrence, structure and RNA sequences of small nuclear RNPs are exhaustively covered in the opening chapter. This is followed by a thorough review of the organization of the major snRNA genes and aspects of their transcription, including the various conserved sequence elements involved. Chapter 3 describes the purification and analysis of snRNP proteins, and culminates in some interesting speculation regarding the possible role of snRNP polypeptides in alternative splicing. The cytoplasmic assembly of small nuclear RNPs and their intracellular transport to the nucleus is concisely and critically

considered in Chapter 4.

The following contribution, the longest in the book, deals with the functions of the abundant U-snRNPs; here the open-minded approach is clearly evident in the evaluation of the data concerning the association of U1 snRNP with the splicing complex, and the special efforts to help the reader through the 'factorology' of the *in vitro* reconstitution studies will be well received. The elegant story of the U7 snRNP and its role in histone pre-mRNA processing is told in Chapter 6, along with that of U11 which seems to be involved in 3' processing of non-histone pre-mRNAs. The snRNP particle RNase P, which matures tRNA molecules by cleaving precursors, is the subject of the next chapter. This review serves to remind us that we do not yet know whether the cleavage/ligation reactions carried out by snRNPs are predominantly catalysed by the RNA or protein components of these particles. Finally, there is an account of some of the successes of analysing yeast snRNAs by exploiting a genetic approach.

There is little overlap between the contributions, and in so much as the book claims to be a review of the literature to 1987, the first seven chapters do this admirably. Chapter 8 on yeast snRNAs, however, reads like the report of a single laboratory. To some extent this is justified because of the particular area covered, but work from other laboratories could have been mentioned in order to perpetuate the feeling of critical, unbiased assessment of current work that characterizes the rest of the book. Could it be that this chapter is more dated than the rest? Nevertheless this is a timely and useful publication, a small and beautiful guide to small nuclear ribonucleoprotein particles. I wonder, however, how big the second edition will be. □

P.C. Turner is a Lecturer in the Department of Biochemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK.

## Reviews on IMMUNOASSAY TECHNOLOGY

Edited by S. B. Pal

Immunoassay techniques are continually being developed and refined, and applied in a broader range of uses. The books in this series comprise review articles and descriptions of methods of immunoassay, which are not reliant on radioisotopes, and reflect the growing importance of immunoassay technologies.

### Volume 1.

October 1988 £37.50 256pp  
234x156mm 0 333 452852

### Volume 2.

October 1988 £37.50 216pp  
234x156mm 0 333 46849X

Order these books through your bookseller or send your order to:

Richard Gedy, Macmillan Press Ltd,  
Houndmills, Basingstoke RG21 2XS,  
UK.

# Degrees of salary and satisfaction

Jim Taylor

*A survey of British graduates six years into their careers shows striking, subject-specific variation in their salaries. But should today's prospective undergraduates take any notice?*

DESPITE its importance to the national economy, information about the graduate labour market in Britain is sketchy and fragmented. Once every ten years, however, the government sanctions a survey of a large sample of graduates from institutions of higher education. The results of the most recent survey (funded by the Department of Employment and the Department of Education and Science, and undertaken in 1986–1987 by Social and Community Planning Research) are now emerging. The *National Survey of 1980 Graduates and Diplomates* provides details on 8,934 anonymous graduates and diplomates during their first six years in the job market (1980–1986).

In this article I look specifically at the data on science graduates, and concentrate on the following questions: How much did graduates earn in 1986? How rapidly did their salaries increase? How many jobs did they hold during their first six years after graduation? What kind of job did they get? What proportion of them were unemployed at the time of the survey? And what proportion obtained further qualifications after their graduation? Answers to all of these questions can

be found for the 15 separate groups of science subjects included in Tables 1 and 2. Information on the sciences is supplemented by information on other subjects to provide a wider perspective.

The most striking result is the variation in salaries between graduates in different subject areas. Average salaries for graduates in mathematics and in computing, for example, were over £16,000 for males and over £13,000 for females in 1986, compared to around £11,000 for both males and females in biochemistry. (Note that average national earnings have increased by about 16% since 1986.) The relatively high salaries of maths and computing graduates were matched only by those of their contemporaries who had read business studies, economics or law. Both groups have benefited from the buoyancy of the job market in financial services and computing. The low proportion of graduates in these subjects who work abroad (Table 1) is further evidence that the home demand for their expertise has been very high during the mid-1980s.

At the other end of the salary scale are graduates in biology and botany. Indeed, in comparison with graduates in business,

economics and law, many of the sciences performed poorly — female graduates in business subjects, for example, had higher average salaries than female graduates in all the sciences (with the single exception of maths and computing), and had higher salaries than males in seven out of the fifteen sciences. Perhaps even more surprisingly, salaries of graduates in the arts, humanities and social sciences were on a par (at least for males) with those of graduates in subjects such as civil engineering, biology, botany, biochemistry, chemistry and geology.

Notably, the salary differential between male and female graduates who worked full-time increased from 9% immediately after graduation in 1980 to 31% six years later. Exactly why women fail to climb the salary ladder as quickly as men is not yet clear, but preliminary findings indicate that about half of this differential can be explained by differences in educational background and qualifications. Male graduates, for example, are far more likely than females to obtain a degree in a science (particularly engineering) than in an arts subject, and science graduates earn more (on average) than arts graduates. ▶

**Table 1** Science graduates in the labour market: comparisons between subjects

| Degree subject<br>(main groups)      | Average salary<br>(full-time workers)<br>in 1986 (£)* |               | Male/female<br>salary<br>differential<br>in 1986 (%) | Real<br>increase<br>in salary<br>1980–1986†(%) |           | Number of<br>jobs held<br>1980–1986 |            | Un-<br>employed<br>in 1986<br>(%) | Working<br>abroad<br>(%) | Qualifica-<br>tions<br>obtained<br>after<br>graduation<br>(%) | Graduates<br>who would<br>do the same<br>or a similar<br>course again<br>(%) | Sample size  |                |
|--------------------------------------|-------------------------------------------------------|---------------|------------------------------------------------------|------------------------------------------------|-----------|-------------------------------------|------------|-----------------------------------|--------------------------|---------------------------------------------------------------|------------------------------------------------------------------------------|--------------|----------------|
|                                      | Males                                                 | Females       |                                                      | Males                                          | Females   | Males                               | Females    |                                   |                          |                                                               |                                                                              | Total        | Females        |
| Pharmacy, pharmacology               | 14,740                                                | 11,230        | 31                                                   | 103                                            | 61        | 2.2                                 | 2.9        | 2.0                               | 4.0                      | 73                                                            | 82                                                                           | 148          | 66             |
| Chemical engineering                 | 14,750                                                | 12,690        | 16                                                   | 63                                             | 57        | 1.5                                 | 2.0        | 1.9                               | 8.8                      | 54                                                            | 80                                                                           | 54           | 5              |
| Civil engineering                    | 12,640                                                | 11,560        | 9                                                    | 68                                             | 89        | 2.1                                 | 1.8        | 1.5                               | 7.8                      | 65                                                            | 63                                                                           | 203          | 10             |
| Electrical engineering               | 15,250                                                | 10,480        | 46                                                   | 80                                             | 48        | 2.0                                 | 2.0        | 0.5                               | 6.5                      | 35                                                            | 88                                                                           | 220          | 3              |
| Mechanical engineering               | 14,630                                                | 10,710        | 37                                                   | 65                                             | 30        | 2.0                                 | 2.1        | 0.6                               | 8.6                      | 51                                                            | 83                                                                           | 162          | 5              |
| Other engineering                    | 15,030                                                | 10,860        | 38                                                   | 79                                             | 56        | 2.1                                 | 2.4        | 1.7                               | 8.5                      | 61                                                            | 79                                                                           | 346          | 16             |
| Agriculture, veterinary              | 13,030                                                | 10,840        | 20                                                   | 70                                             | 55        | 2.0                                 | 2.2        | 2.1                               | 5.0                      | 62                                                            | 77                                                                           | 143          | 44             |
| Biology, botany                      | 11,430                                                | 9,880         | 16                                                   | 53                                             | 35        | 2.2                                 | 2.2        | 3.5                               | 4.7                      | 74                                                            | 74                                                                           | 314          | 110            |
| Zoology, physiology                  | 13,950                                                | 10,240        | 36                                                   | 61                                             | 12        | 2.1                                 | 2.5        | 5.0                               | 8.2                      | 79                                                            | 76                                                                           | 101          | 16             |
| Biochemistry                         | 11,470                                                | 10,850        | 6                                                    | 23                                             | 35        | 2.1                                 | 2.3        | 0.0                               | 9.3                      | 87                                                            | 74                                                                           | 111          | 46             |
| Mathematics, computing               | 16,610                                                | 13,520        | 23                                                   | 111                                            | 80        | 2.0                                 | 2.0        | 2.2                               | 3.9                      | 56                                                            | 85                                                                           | 312          | 83             |
| Physics, maths/physics               | 14,330                                                | 11,540        | 24                                                   | 61                                             | 47        | 1.8                                 | 1.7        | 5.0                               | 6.3                      | 58                                                            | 85                                                                           | 181          | 20             |
| Chemistry                            | 12,800                                                | 11,150        | 15                                                   | 56                                             | 39        | 2.0                                 | 2.0        | 0.0                               | 5.2                      | 71                                                            | 76                                                                           | 204          | 40             |
| Geology, environmental<br>sciences   | 12,850                                                | 10,770        | 19                                                   | 55                                             | 52        | 2.1                                 | 2.4        | 5.7                               | 9.9                      | 77                                                            | 80                                                                           | 141          | 33             |
| Combined sciences                    | 14,460                                                | 11,700        | 24                                                   | 73                                             | 58        | 2.1                                 | 2.3        | 3.3                               | 6.2                      | 70                                                            | 78                                                                           | 391          | 124            |
| <b>Non-sciences</b>                  |                                                       |               |                                                      |                                                |           |                                     |            |                                   |                          |                                                               |                                                                              |              |                |
| Business studies,<br>economics, law‡ | 16,480                                                | 13,420        | 23                                                   | 151                                            | 122       | 2.2                                 | 2.4        | 1.6                               | 2.2                      | 80                                                            | 82                                                                           | 836          | 221            |
| Vocational subjects§                 | 12,650                                                | 9,950         | 27                                                   | 78                                             | 49        | 2.7                                 | 2.6        | 2.4                               | 4.7                      | 76                                                            | 79                                                                           | 212          | 69             |
| All other subjects#                  | 12,640                                                | 10,000        | 26                                                   | 73                                             | 42        | 2.3                                 | 2.6        | 3.7                               | 5.0                      | 72                                                            | 78                                                                           | 3,656        | 1,533          |
| <b>All subjects</b>                  | <b>13,950</b>                                         | <b>10,670</b> | <b>31</b>                                            | <b>81</b>                                      | <b>51</b> | <b>2.2</b>                          | <b>2.5</b> | <b>3.0</b>                        | <b>5.2</b>               | <b>70</b>                                                     | <b>79</b>                                                                    | <b>7,735</b> | <b>(2,537)</b> |

Source: *National Survey of 1980 Graduates and Diplomates* (Employment Market Research Unit, Department of Employment).

\*The mean salary in each subject group overstates the salary which graduates could expect to obtain, because this variable is lognormally distributed. The median salary is approximately £1,300 lower for males and £1,000 lower for females. Mean salary is calculated for full-time workers only. †Allows for consumer price inflation of 47% between 1980 and 1986. ‡Includes accountancy. §Includes architecture and town planning. #Includes education, arts, other social sciences and several combinations.



Table 2 Percentage of graduates in each occupation group by degree subject (sciences only)

| Occupation<br>(main group)* | Average salary<br>(full-time workers)<br>in 1986 (£) |         | Degree subject (main group) |                            |                                       |                                                                |                           |                         |           |                                            |                      |
|-----------------------------|------------------------------------------------------|---------|-----------------------------|----------------------------|---------------------------------------|----------------------------------------------------------------|---------------------------|-------------------------|-----------|--------------------------------------------|----------------------|
|                             | Males                                                | Females | Pharmacy,<br>pharmacology   | Engineering,<br>technology | Agriculture,<br>veterinary<br>science | Biology,<br>botany,<br>zoology,<br>physiology,<br>biochemistry | Mathematics,<br>computing | Mathematics,<br>physics | Chemistry | Geology,<br>environ-<br>mental<br>sciences | Combined<br>sciences |
|                             |                                                      |         |                             |                            |                                       |                                                                |                           |                         |           |                                            |                      |
| Clerical, sales etc†        | 12,280                                               | 9,180   | 1.5                         | 4.0                        | 11.0                                  | 9.8                                                            | 2.9                       | 0.8                     | 6.4       | 4.5                                        | 9.4                  |
| Financial                   | 17,850                                               | 14,360  | 1.1                         | 3.6                        | 4.6                                   | 5.7                                                            | 20.0                      | 7.0                     | 4.3       | 7.2                                        | 13.9                 |
| Computing                   | 14,770                                               | 12,850  | 0.0                         | 5.2                        | 3.4                                   | 6.3                                                            | 38.1                      | 20.9                    | 8.3       | 2.7                                        | 12.3                 |
| Marketing                   | 17,220                                               | 14,070  | 0.0                         | 2.6                        | 6.4                                   | 2.3                                                            | 1.6                       | 2.1                     | 7.0       | 6.9                                        | 6.0                  |
| Administration              | 13,770                                               | 10,880  | 3.6                         | 1.2                        | 1.5                                   | 2.7                                                            | 0.4                       | 3.9                     | 1.9       | 5.4                                        | 3.7                  |
| Higher education            | 11,140                                               | 10,540  | 0.7                         | 1.0                        | 3.4                                   | 3.5                                                            | 4.0                       | 1.5                     | 3.8       | 1.4                                        | 3.0                  |
| Teachers                    | 9,940                                                | 9,480   | 5.8                         | 1.6                        | 5.5                                   | 12.9                                                           | 11.1                      | 8.5                     | 9.1       | 6.9                                        | 12.5                 |
| Welfare workers             | 9,460                                                | 9,420   | 8.4                         | 0.2                        | 18.0                                  | 2.6                                                            | 0.0                       | 0.0                     | 1.9       | 3.4                                        | 3.5                  |
| Health workers‡             | 12,620                                               | 9,720   | 66.5                        | 0.0                        | 0.9                                   | 5.7                                                            | 0.4                       | 0.0                     | 0.6       | 0.7                                        | 2.2                  |
| Scientists                  | 12,180                                               | 10,990  | 8.0                         | 1.7                        | 12.5                                  | 27.6                                                           | 3.9                       | 15.7                    | 37.9      | 32.3                                       | 15.0                 |
| Engineers                   | 14,820                                               | 13,470  | 0.0                         | 61.1                       | 0.0                                   | 2.4                                                            | 13.2                      | 32.2                    | 7.7       | 5.8                                        | 6.8                  |
| Scientific support          | 11,080                                               | 8,300   | 0.7                         | 2.3                        | 0.0                                   | 7.6                                                            | 1.3                       | 1.5                     | 4.7       | 2.4                                        | 3.3                  |
| Building professions        | 12,710                                               | 11,070  | 0.0                         | 5.9                        | 2.7                                   | 0.6                                                            | 0.0                       | 0.0                     | 0.0       | 3.1                                        | 0.4                  |
| Managers                    | 14,100                                               | 11,050  | 3.6                         | 6.5                        | 25.0                                  | 6.6                                                            | 1.3                       | 1.3                     | 4.5       | 8.2                                        | 4.8                  |
| Other                       | —                                                    | —       | 0.1                         | 3.1                        | 5.1                                   | 3.7                                                            | 1.8                       | 4.6                     | 1.9       | 9.1                                        | 3.2                  |
| All occupations             | 13,950                                               | 10,670  | 100                         | 100                        | 100                                   | 100                                                            | 100                       | 100                     | 100       | 100                                        | 100                  |

Source: *National Survey of 1980 Graduates and Diplomates* (Employment Market Research Unit, Department of Employment).

\*A finer disaggregation of occupations is available than the one presented here. †Includes supervisors, personal service workers, foremen, semi-skilled and unskilled workers. ‡Excludes medical and dental professions.

Other factors which affect the male-female salary differential include the type of higher education institution at which males and females take their degree. The proportion of graduates obtaining a Bachelor of Education degree and attending a College of Higher Education, for example, is far higher for females than for males. Not surprisingly, this has a depressing effect on the salaries of female graduates because teachers' salaries are generally low.

That the type of occupation chosen by graduates has a substantial effect on salary prospects can be seen from Table 2. Graduates taking jobs in finance and marketing, for example, can expect to earn far more than those proceeding to a job in welfare work or teaching. This partly explains why graduates in maths, computing and engineering achieve higher salaries than graduates in chemistry, geography, environmental science and biology-related subjects. Thus the low salaries obtained by graduates in biology-related subjects is explained in part by the high proportion who went into low-paid jobs such as teaching. By contrast, 58% of maths and computing graduates were in financial or computing occupations, both of which are very highly paid.

Another result to emerge from the salary data is the remarkable variation between subjects in the rate of increase in graduates' salaries. In maths and computing, for example, the average salary more than doubled in real terms (for males) between 1980 and 1986, compared to an increase of only 23% for male biochemistry graduates. Further investigation reveals that there is no correlation across subjects between starting salary and salary six years later. Obtaining a degree in a subject with a relatively high starting

salary (such as biochemistry) does not therefore guarantee a similar position in the salary league table several years after graduation.

According to the survey, graduates have only a small chance of becoming unemployed compared to the average for all workers. Only 3% of those polled were found to be unemployed against a national unemployment rate of over 11% (in 1986). Although this low rate of unemployment may be due in part to sample bias (unemployed graduates may be more reluctant to respond to such surveys), it is known from other sources, such as the Labour Force Survey, that graduates have a much lower probability of becoming unemployed than do non-graduates. Variations in unemployment between subjects should not be taken too seriously, however, because small changes in the number of unemployed can result in substantial changes in the unemployment rate.

Table 1 shows that graduates change their jobs relatively infrequently, having only about two jobs on average during their first six years in the job market. They therefore tend to be 'stickers' rather than 'movers'. Furthermore, there was little variation in job mobility between graduates in different subjects. Male chemical engineers were the least mobile with 1.5 jobs on average over the six-year period and female pharmacists were the most mobile with 2.9 jobs.

There is, however, considerable variation between the proportions of science graduates obtaining further qualifications after finishing their degrees. The range runs from 35% in electrical engineering to 87% in biochemistry. Yet the proportion of graduates obtaining further qualifications after their graduation is uncorrelated with their salary levels or with the

increase in their salaries during the first six years of their careers. This does not necessarily mean that obtaining further qualifications after graduation is uncorrelated with earnings within individual subject groups.

Finally, it is reassuring to observe that the percentage of graduates who would do the same or a similar course again is remarkably high (over 80% in many cases). Only among civil engineers is there a large proportion who would have chosen to follow a different subject. One would expect there to be a positive correlation between those who would do the same course again and their salary six years after graduation, and that is indeed found to be the case.

What can be concluded from this analysis of part of the *National Survey of 1980 Graduates and Diplomates*? Here I have considered only the salient results, and there is a great deal more to be extracted from the data. But salary prospects for graduates in, for example, biology, botany and biochemistry have clearly been bleaker than for many of their contemporaries who took different subjects, while female graduates have been more poorly paid than males. The survey information is now two years out of date, but it seems unlikely that things have changed radically. It would, however, be unwise to predict what the situation will be in a further four or five years' time, when those now considering which subject to follow at a higher education institution emerge onto the job market. □

*Jim Taylor is in the Department of Economics, University of Lancaster, Lancaster LA1 4YX, UK. His research into the graduate labour market is being partly funded by the Leverhulme Trust.*

# T-cell antigen receptor genes and T-cell recognition

Mark M. Davis\*<sup>†</sup> & Pamela J. Bjorkman<sup>†</sup>

\*The Howard Hughes Medical Institute and <sup>†</sup>Department of Microbiology and Immunology, Stanford University Medical School, Stanford, California 94305-5402, USA

*The four distinct T-cell antigen receptor polypeptides ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ) form two different heterodimers ( $\alpha:\beta$  and  $\gamma:\delta$ ) that are very similar to immunoglobulins in primary sequence, gene organization and modes of rearrangement. Whereas antibodies have both soluble and membrane forms that can bind to antigens alone, T-cell receptors exist only on cell surfaces and recognize antigen fragments only when they are embedded in major histocompatibility complex (MHC) molecules. Patterns of diversity in T-cell receptor genes together with structural features of immunoglobulin and MHC molecules suggest a model for how this recognition might occur. This view of T-cell recognition has implications for how the receptors might be selected in the thymus and how they (and immunoglobulins) may have arisen during evolution.*

THE hallmark of the vertebrate immune system is its ability to mount a highly specific response against virtually any foreign entity, even those never seen before in the course of evolution. In vertebrates, this specificity can be determined by either of two different types of cells known as B or T lymphocytes. These cells are similar morphologically and derive from common haematopoietic stem cell precursors. In the case of B cells, the molecules responsible for specific recognition are the different classes of antibodies that can be expressed either as cell-surface molecules acting as receptors, or in secreted forms serving a variety of purposes including the initiation of complement-mediated killing and the inactivation of viral particles by direct binding. The equivalent recognition molecule on T cells is the membrane-bound T-cell antigen receptor (TCR) which is specific for a combination of foreign antigen with a molecule of the major histocompatibility complex (MHC)<sup>1-3</sup>. One major link between these two different recognition systems is the requirement of most B cells for 'T cell help' in the form of specific growth/differentiation factors (lymphokines) to mount a secreted antibody response<sup>4,5</sup>.

It has been clear for some time that the many different specificities exhibited by antibodies derives from the relatively random juxtaposition of different coding segments (*V*, *D* and *J*) in the genome to produce proteins that differ in their N-terminal domains (*V*, or variable domains) but are the same elsewhere (*C*, or constant)<sup>6</sup>. *V*-region domains from the heavy and light chain polypeptides (*V<sub>H</sub>* + *V<sub>L</sub>*) pair to form the ligand-binding region<sup>7</sup>, and have a range of dissociation constants for antigen from 10<sup>-5</sup> to 10<sup>-14</sup>M (ref. 8). The crystal structures of a number of Fab fragments of antibodies and several antibody:ligand complexes have been solved at the atomic level<sup>9-15</sup>. Thus, the basic architecture of antibodies, and their modes of interaction with antigen are well understood.

By comparison, very little molecular data about TCRs or their ligands has been available until recently. This is due, in part, to the fact that no secretory form of TCRs appears to exist, and serological reagents have been difficult to prepare. Ultimately however, monoclonal antibodies have been derived that can either block or stimulate the activation of helper or cytotoxic T cells, and immunoprecipitate disulphide-linked heterodimers ( $\alpha:\beta$ )<sup>16-20</sup> that are very polymorphic between T cells of different specificity<sup>21-23</sup>, and that are non-covalently associated with as many as five different non-polymorphic polypeptides (the CD3 molecules) thought to be involved in signal transduction<sup>24</sup>. The isolation and analysis of complementary DNA clones encoding these polypeptides<sup>25-30</sup> further indicated that they were antigen receptors because they contained sequences homologous to the *V*, *D* and *J* segments of immunoglobulins, and in T cells these gene segments were found to be rearranged in the vicinity of *C*-like sequences. In addition, transferring the genes for the  $\alpha$

and  $\beta$  chains of the TCR is sufficient to transfer both antigen and MHC specificity between T cells<sup>31,32</sup>.

The complex nature of the TCR ligand has also hindered studies of their function. MHC molecules are polymorphic cell surface glycoproteins that occur in two distinct forms: class I MHC molecules (*H-2* in mice, *HLA-A, B, C* in humans) and class II molecules (*Ia* in mice, *HLA-D* in humans)<sup>33</sup>. It has now been established that one major class of 'objects' that  $\alpha:\beta$  TcR bearing T cells 'see' are peptide fragments of an antigen (presumably derived from intracellular processing) that are bound to MHC molecules at what appears to be a single site<sup>34-41</sup>. The recent completion of the crystal structure of a human class I MHC molecule has identified the probable peptide binding site for this molecule to be a cleft on the outer surface of the molecule, between two  $\alpha$ -helices<sup>40,41</sup>. The structurally homologous class II MHC molecules are predicted to contain a similar peptide binding region<sup>42</sup>.

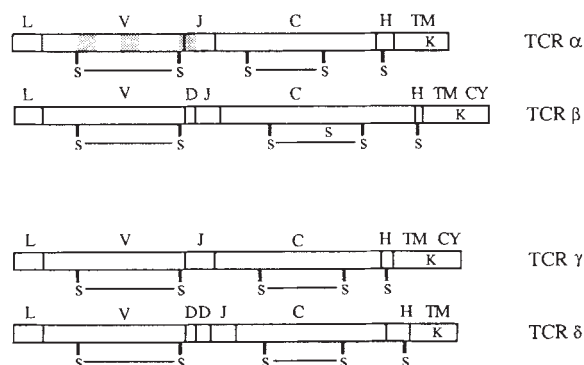
Yet another complicating feature of how T cells see antigen involves the CD4 and CD8 glycoproteins found on the surfaces of T cells. The majority of CD4<sup>+</sup> T cells are of the helper phenotype, whereas most CD8<sup>+</sup> cells are cytotoxic<sup>43</sup>. Although the genes for these proteins do not rearrange, they contain sequences homologous to *V* and (in some cases) *J* at their N-termini, and their expression correlates strictly with the type of MHC molecule with which a T-cell interacts (CD4 with class II MHC and CD8 with class I MHC)<sup>43</sup>. Although not absolutely necessary in TCR  $\alpha:\beta$  dependent recognition<sup>44</sup>, gene transfer studies have shown a marked augmentation of the ability of T cells to respond to the correct target when the appropriate CD4 or CD8 genes are introduced<sup>45-47</sup>.

The purpose of this review is to summarize current knowledge about TCR molecules, their genes, and their patterns of diversity and expression, updating previous reviews on the subject<sup>48,49</sup>. We also attempt to rationalize some of these observations with respect to structural aspects of antigen-MHC recognition and to the evolution of the antigen receptor (TCR and immunoglobulin) genes from possible precursor molecules.

## Structure and expression of TCR polypeptides

The complete primary sequences of the TCR polypeptides were first obtained from cDNA and genomic clones<sup>20-30,50,51</sup>. In most cases, this involved the isolation of cDNA clones expressed in T cells but not in B cells, and then screening these clones for rearrangements in various T-cell DNAs. One surprising result of these efforts was the discovery of a second type of TCR heterodimer ( $\gamma:\delta$ )<sup>50,52,53</sup>, also associated with the CD3 polypeptides, that appears earlier than  $\alpha:\beta$ -bearing cells in ontogeny<sup>54</sup>. These receptors are present on a small percentage of peripheral T cells (1-10%)<sup>52</sup> and on a majority of dendritic T cells in the skin<sup>55,56</sup>. The loci encoding these polypeptides ( $\gamma$  and  $\delta$ ) also

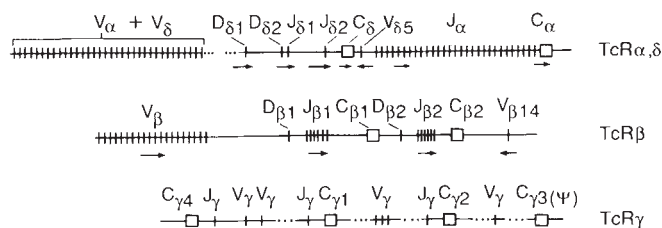




**Fig. 1** Primary sequence relationships of T-cell receptor polypeptides. Leader (L), variable (V), diversity (D) and joining (J) segment contributions are indicated for the variable region domains. The constant portions are divided into the immunoglobulin-like constant (C) domain, a hinge (H) or connecting peptide region that occurs as a separate exon and contains the extra cysteine presumed to form the intrachain bonds, and the transmembrane (TM) and cytoplasmic (CY) portions. An external 'S' indicates the approximate position of a cysteine residue, and an internal 'K' denotes the location of the conserved lysine residue in the transmembrane region. The approximate location of CDR1-, 2- and 3-equivalent regions are indicated by shading in the primary structure of the TCR  $\alpha$ -chain.

have characteristic V, D (in the case of  $\delta$ ), J and C sequences that are rearranged in the appropriate cells<sup>50,51,57,58</sup>. The function of the  $\gamma$ : $\delta$  heterodimer is unknown, whereas TCR  $\alpha$ : $\beta$  appears to be responsible for the specificity of all cytotoxic and helper T cells. The four TCR polypeptides ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ) that form the two types of heterodimers ( $\alpha$ : $\beta$  and  $\gamma$ : $\delta$ ) are all similar to each other, and to immunoglobulins (Fig. 1). Each has V, J and C regions, and  $\beta$  and  $\delta$  also have D regions. The V, J and D elements are very homologous to their counterparts in the immunoglobulins, although the J regions are slightly longer (18–20 as against 13–18 residues). Although the exact number of hypervariable regions in TCR sequences is controversial<sup>59–62</sup>, diversity is clearly evident in the regions equivalent to the three classic immunoglobulin hypervariable regions. In immunoglobulins, these amino acids form the principal points of contact with antigens and are thus referred to as complementarity-determining regions (CDRs)<sup>9,10,63,64</sup>. As is the case for immunoglobulins, the regions equivalent to CDR1 and 2 in TCRs are encoded within the V gene itself, whereas the CDR3-equivalent region is formed by the conjunction of V and J (in TCR  $\alpha$  or  $\gamma$ ), or by V, D and J (in TCR  $\beta$  and  $\delta$ ). The approximate location of the CDR-equivalent regions is indicated for  $V_\alpha$  in Fig. 1.

The constant portions of TCRs are 138–179 amino acids in length, and divisible into four functional regions that often occur as separate exons. The first of these is a domain similar to the immunoglobulin C region of 90–100 amino acids containing characteristic cysteine residues that probably link two  $\beta$ -sheets in the characteristic immunoglobulin-fold. Both the  $C_\beta$  and  $C_\gamma$  sequences are very similar to all immunoglobulin C regions.  $C_\alpha$  and  $C_\delta$  are less similar and have a relatively short distance (50–51 amino acids) between the first two cysteine residues. Nevertheless,  $C_\delta$  (and by extension,  $C_\alpha$ ) is highly similar to the human  $C_{\gamma 1}$  (31%, A. Williams, personal communication) and thus all TCR C regions may fold into immunoglobulin (C-like) domains. An additional cysteine residue resides in a separate exon and is thought to be responsible for the interchain sulphhydryl bond. All four TCR proteins also have distinct hydrophobic regions that appear to be membrane-spanning portions with a highly conserved lysine residue about midway. The positively-charged lysine side chain may be important in interactions with



**Fig. 2** Genomic organization of the T-cell receptor loci in the mouse (not to scale). V, D, J and C region gene segments are indicated and transcriptional orientations (where known) are marked with arrows.

the CD3 polypeptides, four of which are known to have a conserved negative charge (Asp or Glu) in their putative transmembrane regions<sup>65–68</sup>. Nothing equivalent to the CD3 complex has been reported to be associated with membrane immunoglobulin, indicating that the TCR-CD3 complex fulfills some function unique to the T cell.

TCR genes are first rearranged and expressed in the thymus during the earliest stages of T-cell differentiation. TCR  $\beta$ ,  $\gamma$  and  $\delta$ -chain genes are rearranged and transcribed first, followed by the  $\alpha$  chain<sup>51,54,69–71</sup>. The appearance of  $\gamma$ : $\delta$  heterodimers precedes  $\alpha$ : $\beta$  expression by one or two days in fetal ontogeny (in the mouse)<sup>54</sup>. It has been suggested<sup>54</sup> that  $\gamma$ : $\delta$  rearrangement is attempted first and that only those cells that cannot produce a  $\gamma$ : $\delta$  heterodimer (due, for example, to out-of-frame joining of the V and J segments) proceed to  $\alpha$  and  $\beta$  expression. Alternatively, analyses of mice transgenic for a TCR  $\alpha$  chain suggest that these two different types of T cells represent separate compartments and express their respective TCR polypeptides independently (B. Fazekas de St. Groth and L. Berg, unpublished observations). Pronounced differences in TCR  $\gamma$  and  $\delta$  V region sequences between the adult and fetal stage thymocytes<sup>58,72</sup> suggest that these cells have different functions.

## Gene organization and potential diversity

Like the immunoglobulin loci, the TCR genes are divided into arrays of interchangeable coding segments scattered over large tracts of chromosomal DNA (Fig. 2). Specific V, D (in the case of  $\beta$  and  $\delta$ ) and J segments are then fused together to form a complete V-coding domain adjacent to a particular C region. These rearrangements appear to be mediated by sequences similar to those that control the rearrangement of immunoglobulin gene segments (conserved heptamer and nonamer sequences spaced by 12 or 23 nucleotides)<sup>73,74</sup>. Thus, the process of recombining gene elements appears very similar for TCRs and immunoglobulins, with both looping-out and deletion<sup>75–78</sup> and inversion<sup>79,80</sup> occurring, depending on the orientation of the sequences being joined. Also reminiscent of immunoglobulin heavy-chain genes<sup>81</sup> is the random addition of nucleotides between the V, D and J gene segments (N region addition) that can create additional diversity (see below).

The TCR  $\beta$  locus is the first large antigen-receptor locus to be completely mapped. It spans 700–800 kilobases of the chromosome in most strains of mice<sup>82,83</sup>. Two nearly identical (and functionally equivalent) C regions are tandemly arranged with one D element and a cluster of six J elements 5' to each<sup>48,49</sup>. One  $V_\beta$  element is located 3' to the DJC elements and rearranges by inversion<sup>79</sup>. The other regions, some 20 to 30 in number<sup>60,84</sup>, are 5' to DJC, and appear to rearrange by the looping out and deletion of the intervening DNA<sup>77,78,82,83</sup>, although evidence for sister chromatid exchange has also been obtained<sup>48</sup>. Strains of mice have been identified that have deleted a large fraction ( $\approx 10$ ) of these V segments<sup>85</sup> or one of the  $J_\beta$  clusters<sup>86,87</sup> yet appear immunologically normal.

The TCR  $\gamma$  locus is less diverse (as indicated in Fig. 1) with three different functional  $J_\gamma$ - $C_\gamma$  loci in most mouse strains<sup>57,88</sup>

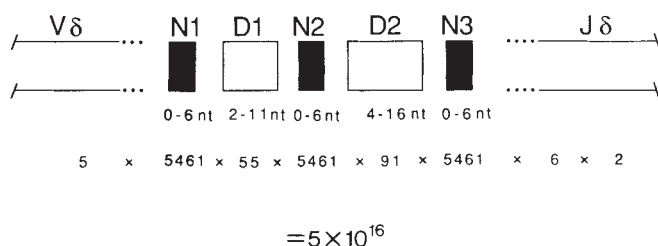
**Table 1** Sequence diversity in T-cell receptor and immunoglobulin genes

|                              | Immunoglobulin |          | TCR $\alpha:\beta$ |                | TCR $\gamma:\delta$ |                                                                          |
|------------------------------|----------------|----------|--------------------|----------------|---------------------|--------------------------------------------------------------------------|
|                              | H              | $\kappa$ | $\alpha$           | $\beta$        | $\nu$               | $\delta$                                                                 |
| Variable segments            | 250-1,000      | 250      | 100                | 25             | 7                   | 10                                                                       |
| Diversity segments           | 10             | 0        | 0                  | 2              | 0                   | 2*                                                                       |
| Ds read in all frames        | Rarely         | —        | —                  | Often          | —                   | Often                                                                    |
| N-region addition            | V-D, D-J       | None     | V-J                | V-D, D-J       | V-J                 | V-D <sub>1</sub> , D <sub>1</sub> -D <sub>2</sub> ,<br>D <sub>1</sub> -J |
| Joining segments             | 4              | 4        | 50                 | 12             | 2                   | 2                                                                        |
| Variable region combinations | 62,500-250,000 |          |                    | 2,500          |                     | 70                                                                       |
| Junctional combinations      | $\sim 10^{11}$ |          |                    | $\sim 10^{15}$ |                     | $\sim 10^{18}$                                                           |

Calculated potential amino acid sequence diversity in T-cell receptor and immunoglobulin genes without allowance for somatic mutation. The approximate number of V gene segments are listed for TCR  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  contrasted with  $V_H$  and  $V_L$ . The first two hypervariable regions of immunoglobulins (CDR1 and CDR2) and their equivalents in TCRs are encoded within the V gene segments. The pairing of random V regions ( $V_H \times V_L$ ,  $V_\alpha \times V_\beta$  or  $V_\gamma \times V_\delta$ ) generates the combinatorial diversity listed as 'variable region combinations'. The magnitude of combinatorial diversity in TCRs is lower than in immunoglobulins because of the decreased number of TCR V gene segments. Estimates for the number of amino acid sequences that might result from diversity within the junctional region are contrasted for TCRs and immunoglobulins. The third immunoglobulin hypervariable region (CDR3) and its TCR equivalent are encoded almost entirely within the D and/or J region gene segments. The last few amino acids encoded by a TCR V-gene segment can contribute to the TCR CDR3-equivalent region, but the effects of these residues on junctional diversity are not included in these calculations. The mechanisms for diversity generation within the junctional region used for this calculation include usage of different D and J gene segments, N region addition up to six nucleotides at each junction, variability in the 3' joining position in V and J gene segments, and translation of D regions in different reading frames. Numbers are corrected for out-of-frame joining codon redundancy and N-region mimicry of germline sequences as detailed in Elliott *et al.*<sup>72</sup>.

Only seven V gene segments have been identified to date<sup>57,88-90</sup>. It is not known if the different  $C_\gamma$  sequences are functionally equivalent. In humans, the TCR  $\gamma$  segments occupy 150 kilobases (kb) of chromosomal DNA<sup>91</sup>.

The TCR  $\alpha$ ,  $\delta$  locus is unique in that many of the TCR  $\delta$  coding segments are located entirely within the V and J segments of the  $\alpha$  gene<sup>58,72</sup> (Fig. 2) and are deleted in most  $\alpha:\beta$ -bearing T-cell lines and hybrids<sup>92</sup>. Although there is some overlap in V region usage, the V gene segments used for  $\alpha$  and  $\delta$  chains are largely distinct<sup>72</sup>. An estimated 75-100 V regions could be used by the TCR  $\alpha$  chain,<sup>48,62</sup> but only about 10 V segments have been identified for the  $\delta$  chain and these are split between fetal and adult usage<sup>58,72</sup>. Although a number of the fetal  $V_\delta$  coding regions may be located 3' to  $V_\alpha$  sequences, other  $V_\delta$ s appear interspersed with  $V_\alpha$ s<sup>72</sup>. Similar to the case of the TCR  $\beta$  locus, a  $V_\delta$  gene segment has recently been found just 3' to  $C_\delta$  and in the opposite transcriptional orientation, indicating that TCR  $\delta$  also uses both looping out and deletion and inversion in its rearrangements<sup>80</sup>. Another interesting feature of this locus is that the TCR  $\alpha$  gene greatly exceeds  $\delta$  (or any other antigen receptor gene) in the large (>50) number of J region segments<sup>61</sup>. Two D and two J gene segments are evident in TCR  $\delta$ <sup>58</sup>.



**Fig. 3** Nucleotide sequence diversity in the junctional region of the adult TCR  $\delta$  chain. The relative contribution to diversity of the different V, N, D or J elements are schematized. Underneath each N and D segment is the number of nucleotides (nt) possible for each element, and below that are the number of permutations this number of nucleotides allows assuming random nucleotide usage. The numbers under the  $V_\delta$  and  $J_\delta$  gene contributions indicate the variation in the joining position for those gene segments (or junctional diversity). At the extreme right, the factor of 2 is included to account for the two different  $J_\delta$  gene segments.

Although the basic organization of the TCR loci is very similar to that of the immunoglobulins, a closer look at the generation of diversity in TCR heterodimers (Table 1) indicates a striking concentration of sequence polymorphism in the CDR3-equivalent region. As summarized in this table, TCR genes generally employ many fewer V gene segments than do immunoglobulins, from as many as 100  $V_\alpha$  sequences to the seven or so  $V_\gamma$  sequences (versus the 250-1000  $V_H$  or  $V_\kappa$  gene segments). Assuming that the V domains of TCRs pair to form a combining site (as is the case for immunoglobulin V regions; see next section), the combinatorial diversity (resulting from random pairing of V domains; for example  $V_\alpha \times V_\beta$ ) in these antigen receptor molecules shows an even greater disparity when compared to immunoglobulins. In contrast, the diversity at the V-J junction of TCRs greatly exceeds that of immunoglobulins, both in TCR  $\alpha:\beta$  and TCR  $\gamma:\delta$ . Several unique features of TCR genes contribute to diversity in this region:

- (1) N-region diversification has now been documented in all four TCR polypeptides<sup>72,90,93-95</sup> but only occurs in the heavy chain locus of immunoglobulins<sup>81</sup>. This is one of the most significant components in the increased junctional diversity in TCRs.
- (2) As noted previously<sup>48</sup>, the large number of TCR J region gene segments ( $\approx 50 J_\alpha$ s and 12  $J_\beta$ s), compared to the four Js found in mouse immunoglobulin heavy and  $\kappa$  chain loci<sup>6</sup> is a major factor in the increased junctional diversity in  $\alpha:\beta$  TCRs compared to immunoglobulins.
- (3) Several  $V_\alpha$ s,  $V_\gamma$ s and  $V_\delta$ s<sup>58,72,90,96</sup> have been shown to be flexible with respect to the position of their 3' joining points, a phenomenon that has not been seen in immunoglobulins.
- (4) The ability to use both D regions in the adult  $\delta$  chain<sup>58,72</sup> and the fact that N region addition occurs at three different positions greatly increases the available diversity. (See Fig. 3 and ref. 72.) In addition, the translation of D region sequences in all three possible reading frames is rare in immunoglobulin heavy chains but common in TCR  $\beta$ <sup>97</sup> and  $\delta$ <sup>72</sup> chains, compensating for the smaller number of TCR germline segments (see Table 1).

We do not include the effects of somatic hypermutation on diversity as this appears to be specific to immunoglobulins. The inclusion of this apparently random mutational mechanism would have the effect of increasing both junctional and  $V \times V$  combinatorial diversity for those molecules.

The most striking illustration of the skewing of diversity in



TCR polypeptides towards the junctional region is the diversity of the  $\delta$ -chain gene (Fig. 3). Only a small number (5–10) of  $V$  gene segments seem to be used at any given stage, yet in adult thymocytes the junctional region has contributions from both  $D$  regions, three  $N$  regions and variability in the joining position for both  $V$  and  $J$ . It has been estimated that  $10^{13}$  possible amino acid sequences could be derived from the junctional region alone<sup>72</sup>. This is an enormous number of different sequences, particularly in a region that encodes only 10–20 amino acids. We believe that this concentration of diversity in the  $V$ - $J$  junction of  $\delta$  and the other TCRs, together with the comparative lack of diversity elsewhere, is a specific consequence of what TCRs (the  $\alpha:\beta$  type, at least) are known to do; namely recognize very diverse small molecules (such as peptides) embedded in much less diverse and larger MHC molecules. Structural features of MHC and immunoglobulin molecules support this contention and are discussed in the following section.

### Immunoglobulin and MHC structures

Sequence data suggest that TCR variable regions are folded into  $\beta$ -sheet structures resembling immunoglobulin variable regions<sup>59–62,98,99</sup>. In antibodies, the variable regions from the heavy and light chains ( $V_H$  and  $V_L$ ) are paired such that the three complementarity-determining regions (CDR1, 2, and 3) from each domain are clustered at the ends of the Fab arms of the molecule, forming the antigen-binding site<sup>9,10,100</sup>. Many of the residues in the  $V_L$ - $V_H$  interface are conserved, as is the overall geometry of  $V_L$ - $V_H$  pairing<sup>101,102</sup>, resulting in a similar arrangement of the CDRs around the binding sites of the molecules whose structures are known. The contact surface between  $V_L$  and  $V_H$  is unique among proteins with  $\beta$ -sheet interfaces<sup>101</sup>, and most of the amino acids that seem important in these contacts (in immunoglobulin) are identically placed in TCR  $V$  region sequences<sup>62,98,99</sup>. Thus it seems reasonable to assume that TCR and immunoglobulin  $V$  domains will fold in the same fashion, and that the chain pairing and resulting combining sites of  $V_\alpha:V_\beta$  and  $V_\gamma:V_\delta$  will also be similar to those of  $V_H:V_L$  combinations<sup>99</sup>.

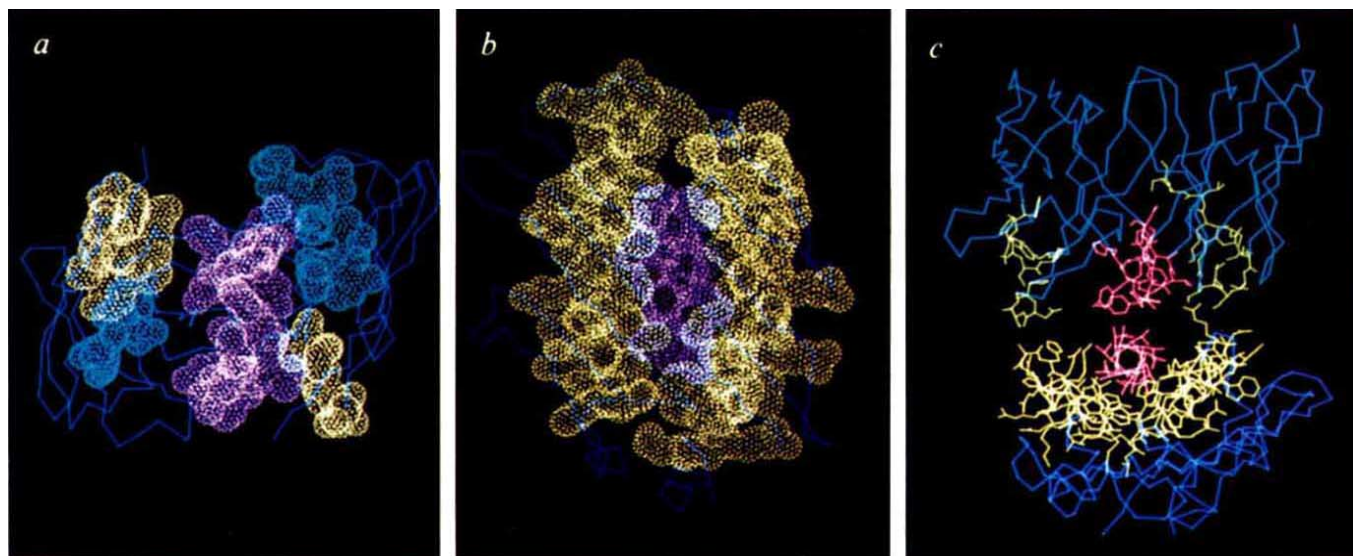
When the combining site of an immunoglobulin molecule is viewed from the direction of the antigen (Fig. 4a), the first and second CDRs on  $V_L$  are separate from their counterparts on  $V_H$ , and the space between them is occupied by the two CDR3 regions. Most of the residues at the  $V$ - $J$  junction and at the beginning of the  $J$  region of  $V_L$  are on the surface of the combining site in positions to interact with a bound antigen. This is also true for most residues at the  $V$ - $D$  and  $D$ - $J$  junctions of  $V_H$ , and the  $D$  region and the first few  $J$  region residues. Thus, the diverse residues in the  $V$ - $J$  junctional (CDR3) region are located next to each other at the center of the combining site. Structural studies of hapten-Fab complexes<sup>14,15</sup> showed that the binding site for hapten is a cleft located between the third CDRs of  $V_H$  and  $V_L$ . More recent structures of Fab's complexed to protein antigens<sup>11–13</sup> have demonstrated that this cleft is not a major feature of the (protein) antigen-antibody interface. Instead, side chains from all six CDRs contact the antigen, so that the entire interface is a rather flat surface with protrusions and depressions in the antibody complementary to the antigen surface.

A TCR presumably used the diversity in its hypervariable regions to recognize an MHC-peptide complex. The structure of a class I histocompatibility molecule suggests that the binding site for peptide is on the top surface of the molecule, located between the  $\alpha$ -helical regions of the polymorphic  $\alpha_1$  and  $\alpha_2$  domains<sup>40,41</sup>. If this site were occupied by an antigenic peptide, the top surface of the MHC-peptide complex would be quite flat (as indeed is the case for HLA-A2 which co-crystallized with unknown molecule(s) in the peptide binding site)<sup>40</sup>. Fig. 4b shows the top surface of HLA-A2 with a (hypothetical) 12-amino-acid peptide bound. A comparison of Fig. 4a with 4b shows that the MHC  $\alpha$ -helices are separated by about the same

distance ( $\approx 18$  Å) as the CDR1 and CDR2 loops of  $V_H$  are separated from those of  $V_L$ . Thus, the flat surface of an MHC protein-peptide complex could interact with the combining site of a TCR (assuming it is immunoglobulin-like) such that the CDR1- and CDR2-equivalent regions on  $V_\alpha$  and  $V_\beta$  contact the side chains of the MHC  $\alpha$ -helices, leaving the centrally-located CDR3-equivalent regions contacting the bound peptide. In Fig. 4c, the  $V_H$ - $V_L$  combining site (here representing a TCR combining site) is aligned over a side view of the MHC-peptide complex. The figure demonstrates the complementarity between the two surfaces, and the alignment of CDR1 and CDR2 with MHC determinants and of CDR3 with peptide determinants. Alternatively, the TCR could bind to the MHC molecule at some angle (for example  $90^\circ$ ) from that shown in Fig. 4c such that the CDR3-equivalent regions would straddle the peptide-binding groove of the MHC molecule. In this case, the CDR3-equivalent regions would contact parts of the two MHC  $\alpha$ -helices and a portion (but not all) of the bound peptide. The CDR1- and CDR2-equivalent regions of the  $\alpha$  and  $\beta$  chains would also contact parts of the bound peptide and the MHC  $\alpha$ -helices. This alignment allows the TCR junctional residues to have some role in peptide binding, but is less appealing because the limited diversity in the TCR CDR1- and CDR2-equivalent regions would be required to interact with a large number of peptide antigens. Another reason we favour the alignment shown in Fig. 4c is that it would tend to explain how a particular TCR  $V$  region might correlate strongly with reactivity to a given MHC molecule (as has been seen recently for a  $V_\beta$  sequence)<sup>103</sup>, as this configuration allows the most contact between residues encoded with the  $V$  gene segment and an MHC molecule.

This model for TCR and MHC-peptide interactions is also consistent with recent work correlating TCR  $\alpha:\beta$  sequences with known antigen-MHC specificities. In particular, Fink *et al.*<sup>104</sup> and Winto *et al.*<sup>96</sup> noted that mouse TCRs specific for the C-terminal peptide of cytochrome *c* preferentially use a single  $V_\alpha$  gene segment. Only mice expressing I-E class II MHC molecules respond to cytochrome *c*<sup>105</sup>. As the  $\alpha$  subunit of I-E class II antigens is essentially non-polymorphic<sup>106</sup>, our model for TCR/MHC-peptide interaction suggests that, in the majority of cytochrome *c*-reactive T-cell clones, the CDR1- and CDR2-equivalent regions encoded within the  $V_\alpha$  gene segment would interact with a region of the non-polymorphic class II  $\alpha_1$  domain, leaving the CDR1- and CDR2-equivalent regions encoded by the TCR  $V_\beta$  gene segment to interact with the polymorphic I-E  $\beta$  chain. Indeed, among cytochrome *c*-specific TCRs that use this  $V_\alpha$  gene segment, a correlation has been found relating  $V_\beta$  gene segment use and reactivity with different I-E  $\beta$  chains<sup>104,107</sup>. Furthermore, both Fink *et al.*<sup>104</sup> and Winto *et al.*<sup>96</sup> found that in some cytochrome *c*-specific T-cell clones, changes limited to the junctional regions of TCR sequences altered the specificity for peptide without altering MHC specificity. In addition, the finding that some cytochrome *c*-specific TCRs show selection for certain amino acids with the CDR3-equivalent region<sup>108</sup> suggests that these residues are important for peptide recognition. In another system, analyses of myelin basic protein-recognizing T-cell clones has revealed a striking conservation of amino acid sequence in the  $D_\beta$  region, although the  $D_\beta$  nucleotide sequence differs between the clones<sup>109</sup>.

Even within the confines of this model for TCR/MHC-peptide interaction, one would not always expect a direct correlation of junctional residues with restriction to a particular peptide, because changes in MHC residues could dramatically affect the conformation of a bound peptide, and possibly also its orientation in the peptide-binding site. Also, the surface of a peptide-MHC complex is large enough to allow a TCR (assuming a binding site similar to that of an immunoglobulin) to bind in different registers along the MHC  $\alpha$ -helices (see Fig. 4c, legend). Thus, the  $V_\alpha$  and  $V_\beta$  gene segments used by T cells restricted to the same MHC molecule might be very different. Simple correlations between TCR hypervariable region sequences and



**Fig. 4** Representations of the structures of *a*, the immunoglobulin combining site; *b*, the peptide-binding site of an MHC molecule; and *c*, the alignment of CDRs in a hypothetical TCR over a peptide-MHC complex. *a*, The arrangement of CDRs in an immunoglobulin antigen-binding site (Fab J539)<sup>122</sup> viewed from above (from the position of the antigen). The carbon-α backbone of  $V_H$  and  $V_L$  is shown in pink with van der Waals surfaces highlighting the three CDRs from each domain (CDR1: blue, CDR2: pink, CDR3: green). The first and second CDRs from one variable domain (for example, immunoglobulin  $V_L$  or TCR  $V_\alpha$  or  $V_\gamma$ ) are separate from their counterparts on the partner variable domain (immunoglobulin  $V_H$  or TCR  $V_\beta$  or  $V_\delta$ ). The space between them is occupied by the third CDR from each V domain. The similarities between immunoglobulins and TCRs suggests that TCRs may have a combining site that preserves these same general features (see text for details). *b*, Top surface of an MHC molecule with a (hypothetical) bound peptide. The carbon-α backbone of the  $\alpha_1$  and  $\alpha_2$  domains of HLA-A2 (ref. 40) is shown in pink with van der Waals surfaces highlighting the two  $\alpha$ -helices (blue). Van der Waals surface of a hypothetical bound peptide (a 12-mer polyvaline  $\alpha$ -helix) that has been fitted between the HLA  $\alpha$ -helices is shown in green. The putative recognition site for processed foreign antigens is located on the top surface of the molecule between two  $\alpha$ -helices for the human class I histocompatibility molecule HLA-A2<sup>40,41</sup>. The N-terminal  $\alpha_1$  and  $\beta_1$  domains of class II MHC molecules are predicted to have a similar tertiary structure and peptide binding site<sup>42</sup>. Note that the distance between the MHC  $\alpha$ -helices, and the separation of the first and second CDRs of each V-region in the combining site shown in part *a* are very similar. (Figures are to scale with respect to each other). *c*, Model for TCR interaction with a peptide-MHC complex. The combining site CDRs (*a*) are shown aligned over the peptide-MHC complex (*b*). The molecules in this figure are rotated  $\sim 90^\circ$  with respect to their orientations in (*a*) and (*b*). V domains of Fab J539 (ref. 122; top of figure) here represent the V domains of a TCR bound to a peptide-MHC complex (bottom of figure). The carbon-α backbone of  $V_L$  and  $V_H$  (pink) is shown with main and side-chain atoms of CDR1 and CDR2 (blue) from each domain. The main and side-chain atoms of the CDR3s of each domain are shown in green. The carbon-α backbone of the  $\alpha_1$  and  $\alpha_2$  domains of HLA-A2 (pink) is shown with the main and side-chain atoms of the two  $\alpha$ -helices in blue and the main and side-chain atoms of a hypothetical peptide bound between the two helices in green. The relatively flat surface of the V-region combining site is complementary to the flat surface of the MHC-peptide complex, with CDR1 and CDR2 of each V domain 'fitting' over an MHC  $\alpha$ -helix. The CDR3s from each V domain are then aligned over the peptide (see text for details). Because the antibody (and presumably TCR) V regions pair with approximate two-fold (dyad) symmetry<sup>9,10</sup>, a similar interaction of CDR1 and CDR2 with the MHC  $\alpha$ -helices and CDR3 with peptide would be accomplished if the  $V_H$ - $V_L$  dimer shown in this figure were rotated by  $180^\circ$  about the pseudo-dyad axis between the two V regions (axis is vertical in this figure). Note that the relative sizes of the combining site (*a*) and the peptide-binding site of the MHC molecule (*b*) would allow TCRs to be bound in different registers along the MHC  $\alpha$ -helices, depending on the particular peptide bound to that MHC protein. Thus the V genes employed by different T cells restricted to the same MHC molecule would not be expected to be identical. Furthermore, the C-terminal few amino acids encoded by a V gene are part of the V-J or V-D-J junction, and would be predicted by this model to interact primarily with peptide rather than MHC determinants.

MHC and peptide specificities may never be possible due to the spatial proximity of CDR residues in the combining site. Nevertheless, this model for TCR/MHC-peptide interaction may provide a useful framework for the design of experiments to test which specific TCR residues affect MHC or peptide recognition, and has implications for the evolution of the T and B cell recognition systems.

Because adult  $\gamma$ : $\delta$  TCRs exhibit the strongest dichotomy with respect to the spectrum of V region diversity compared with junctional diversity (Table 1), we suggest, based on this model, that these receptors also see peptides on MHC-like molecules, although they may be more restricted in the MHC molecules with which they interact. Recently, a  $\gamma$ : $\delta$ -bearing T-cell line has been shown to exhibit MHC-linked recognition specificity<sup>110</sup>, supporting the view that  $\alpha$ : $\beta$  and  $\gamma$ : $\delta$  TCR heterodimers may have similar recognition properties.

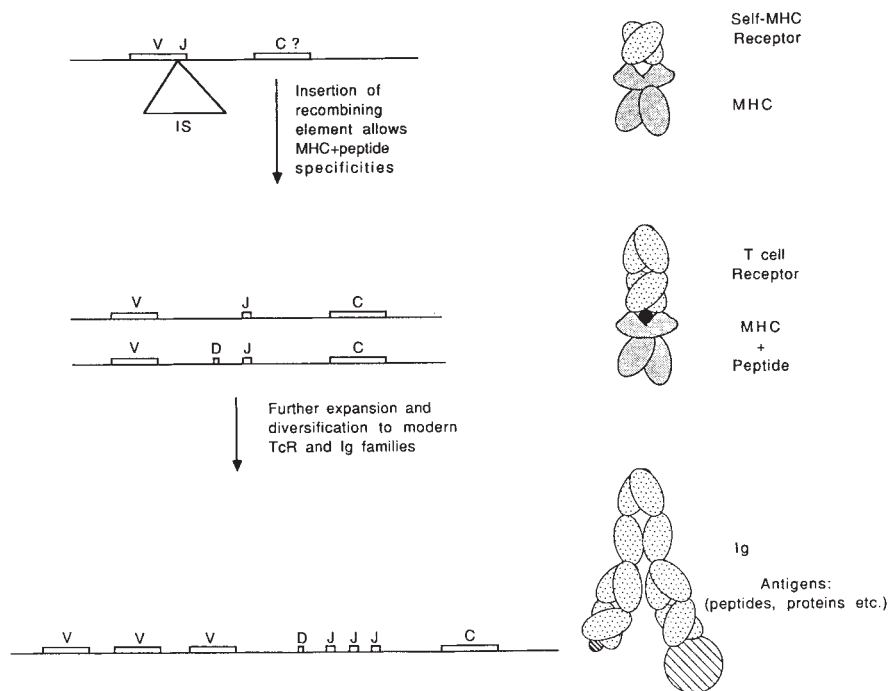
## Evolutionary implications

In addition to its possible usefulness in understanding TCR:ligand interactions, this model can also provide a

framework for thought about the possible evolution of antigen-receptor genes. In particular, it provides a crucial rationale for the 'capture' of a transposable element, which appears to have been the major step in the transition from a static gene to a rearranging gene. Following the original suggestion of Tonegawa and colleagues<sup>111</sup>, we envisage a situation in which a primordial VJ domain acquires an insertion element and thus the ability to join V and J by excision. (A variation of this, resulting in a V-D-J-C gene, has also been suggested by Hood *et al.*<sup>112</sup>.) What has been more difficult to explain is how such an event could become useful to an organism in a reasonable evolutionary time frame, as many examples of transposon insertion are known which merely inactivate the target gene. We suggest that the immediate precursor to the T-cell antigen receptor was a V region heterodimer or homodimer that served as a self-MHC receptor (illustrated in Fig. 5). The acquisition of a transposable element between the V and J sequences of such a molecule would create junctional diversity at that site and thus allow it to bind to different antigens embedded in the original target MHC molecule. Thus, a single genetic event could create a



**Fig. 5** Possible evolutionary progression in the evolution of T- and B-cell antigen receptor genes. A primordial anti-MHC receptor expressed as a V region homodimer or heterodimer suffers an insertion (IS) in its VJ exon by a transposable element (top left). Excision of this element (V-J joining) or elements (V-D-J joining) creates junctional diversity and the ability to recognize a variety of peptides on MHC molecules. This primordial MHC receptor may or may not have had immunoglobulin-like C region domains. Subsequent gene segment duplication and divergence might then have resulted in the modern antigen receptor loci (TCRs; middle left, and immunoglobulins, bottom left). An MHC molecule is represented schematically (top right) in dark shading with the two V-like domains of the primordial anti-MHC receptor (light shading) aligned over the peptide-binding site on the MHC molecule. In this case, the MHC molecule is depicted without a bound peptide, as interactions with a peptide are assumed to be irrelevant. A TCR (light shading; middle right) is represented interacting with an MHC molecule (dark shading) and an embedded peptide (small dark circle) (see Fig. 4c). We suggest that the TCR may have a combining site for antigen and MHC that is structurally similar to that of an antibody (bottom right). The antibody is capable of binding both peptide (small circle) and protein (large circle) antigens in the absence of an MHC molecule. For illustrative purposes, the immunoglobulin molecule is shown binding two different ligands. In reality, because the two Fab arms of an antibody are identical, both would bind the same ligand.



primitive T-cell receptor gene, and subsequent gene duplication and divergence could result in both the 'modern' TCRs able to recognize many peptides on many MHC molecules, and antibodies, which have diverged (and diversified) their V regions to the point that they are able to bind to non-MHC molecules.

We suggest the CD8 heavy-chain gene (Lyt 3 in the mouse) as the nearest relative of the precursor of TCR and immunoglobulin genes because of the striking similarity of its V- and J-like sequences to those of the antigen receptor genes<sup>113,114</sup>. The CD8 heavy chain forms a heterodimer with the light chain (Lyt 2 in the mouse), which also has an N-terminal V domain and a partial J region sequence. Both the heavy-light heterodimer and the light chain homodimer of CD8 are effective in augmenting T-cell recognition of class I MHC determinants (J. Parnes, personal communication and refs 45, 47). It is likely that the homodimer and heterodimer forms of CD8 pair with a similar geometry, by analogy to immunoglobulin light-chain homodimers that are known to pair with the same approximate geometry as heterodimers<sup>100</sup>. The CD8 heavy chain and light chain<sup>98</sup> each conserve the amino acids that are important in the  $V_H-V_L$  interface<sup>101,102</sup>, suggesting that both forms of CD8 contain an immunoglobulin-like combining site. CD4, which appears to be a monomer, has both an N-terminal V and partial J sequences and another more weakly homologous V-like sequence in the middle of the rather long polypeptide<sup>116</sup>. It may therefore be able to form an intramolecular V-like heterodimer. We suggest that CD8 (and perhaps CD4) bind to MHC molecules in basically the same way as suggested for TCR  $\alpha:\beta$  and  $\gamma:\delta$  V-region pairs (Fig. 5). By necessity, this implies that TCRs and CD4 or CD8 could not bind simultaneously to the same MHC molecule. This contention is supported by the work of Mescher and Goldstein<sup>117</sup> who have shown that class I molecules that are irrelevant to the specificity of an alloreactive T cell can nonetheless augment the T-cell response, suggesting that CD8 and TCR can interact with different MHC molecules.

One further indication of the evolutionary (and thus, perhaps functional) relatedness of the antigen-receptor genes and the CD8 genes is the fact that the CD8 and immunoglobulin  $\kappa$  genes are closely linked in both the mouse and human genomes. In

the mouse, the CD4 and the TCR  $\beta$ -chain loci are also located on the same chromosome (six) although at a considerable distance<sup>43,48</sup>. The linkage of these genes suggests a common ancestry followed by a step-wise series of duplications and divergences consistent with the scheme illustrated in Fig. 5.

### Implications for thymic selection

One of the central questions in T-cell biology is how and why T cells emerge from the thymus with the ability to recognize foreign peptides on self-MHC molecules but to a lesser extent or not at all on other MHC alleles. Although this issue is still controversial, a number of experiments have suggested that specific T cells are preselected in some fashion for their compatibility with self-MHC in the absence of foreign antigens<sup>118</sup>. This model is referred to as 'positive' thymic selection and ensures that the TCRs in any given individual have an affinity for the MHC alleles present in that individual's thymus. 'Negative' selection in the thymus refers to the removal of self-reactive T cells to ensure that individuals are tolerant to their own proteins<sup>119</sup>.

One prediction of our model of T-cell recognition is that positive selection and negative selection might be based largely on  $V_\alpha:V_\beta$  pairings (and perhaps  $V_\gamma:V_\delta$ ) interacting with specific MHC molecules, irrespective of their V-J junctions. The structural complementarity we propose between the TCR CDR1- and CDR2-equivalent regions and the  $\alpha$ -helices on an MHC molecule (depicted using immunoglobulin V-regions and HLA-A2 in Fig. 4c) may also provide a molecular explanation for the original suggestion of Jerne that antigen receptors and MHC molecules coevolved to have some affinity for each other<sup>120</sup>. Such a requirement for selecting  $V_\alpha:V_\beta$  pairs, rather than  $V_\alpha$  or  $V_\beta$  genes separately, would also explain why there is, in general, no correlation between MHC type alone and the expression of particular  $V_\alpha$  or  $V_\beta$  gene segments<sup>121</sup>.

### Deletion of specific $V_\beta$ sequences

Recent studies<sup>123-125</sup> have indicated that the percentage of T cells expressing particular  $V_\beta$  elements is significantly lower in mice that combine one of the *Mls* loci with specific MHC

haplotypes. The deletion of  $V_{\beta 17a}$ -expressing T cells in I-E-bearing mice, originally thought to result from a high affinity of  $V_{\beta 17a}$  for I-E itself, may also depend on cell-type-specific antigen(s)<sup>126</sup>. These observations show that T cells can be deleted without regard to either  $V_{\alpha}$  expression or junctional sequences in the TCR  $\alpha$  or  $\beta$  chain. This could be interpreted as suggesting that specific germline  $V_{\beta}$  encoded residues might be directly involved in antigen (that is, peptide) contact, although the nominal antigens have not been identified in any of these examples. To reconcile these data with our model for TCR and MHC-peptide interactions we suggest the following:

- (1) The *Mls* loci may not encode molecules that appear as peptides bound to the MHC, but could instead be responsible for post-translational modifications to MHC molecules themselves or an alteration in the function of MHC-bearing antigen presenting cells<sup>127,128</sup>, resulting in the increased affinity of certain TCRs for the MHC portion of an *Mls*-modified MHC-peptide complex.
- (2) Some antigens (the putative cell-type-specific antigen(s) involved in the deletion of  $V_{\beta 17a}$ -expressing T cells, or the *Mls* antigen(s), for example) could cause conformational changes in the MHC molecules that bind them, inducing a tighter than normal 'fit' of the MHC molecule with specific TCR  $V_{\beta}$ s. In the above-mentioned cases, the MHC molecule that seems most involved is I-E, which is only polymorphic in its  $\beta$  subunit<sup>105</sup>. We would argue that the TCR  $V_{\beta}$  molecules involved in the clonal deletion of T cells would principally contact the surface of the  $E_{\alpha}$  portion of the I-E heterodimer.
- (3) Regions of hypervariability other than those corresponding to the classical CDR1, 2 and 3 portions of immunoglobulins have been reported in analyses of  $V_{\beta}$  sequences<sup>59,62</sup>. One of the most prominent of these regions (residues 62 and 71) is very close to the combining site in an antibody molecule (although not in the plane of the nearly flat surface defining the antibody interface with antigen). It is possible that a 'long' MHC-binding antigen could bind such that one or both of its ends reached out of the peptide binding site, allowing contact between it and this other hypervariable part of the TCR  $V_{\beta}$ .
- (4) Some germline-encoded  $V_{\beta}$  residues could contact antigen as well as MHC determinants if TCRs recognize peptide-MHC complexes as shown in Fig. 4c, because of the close proximity of CDR residues in the combining site (represented by an antibody combining site in Fig. 4b). Alternatively, some subset of TCRs (such as those under discussion) may recognize certain antigen-MHC complexes at some angle other than that depicted in Fig. 4c such that germline encoded  $V_{\beta}$  residues contact antigen as well as MHC determinants.

Any of these mechanisms could allow a sufficiently tight interaction (leading to clonal deletion of T cells) between the  $V_{\beta}$  part of a TCR heterodimer and the presumptive antigen-MHC complexes under discussion.

## Summary and prospects

Much has been learned recently about TCR genes and their products, and their similarities to immunoglobulins are immediately apparent. The current challenge is to further define how and why the T-cell recognition system differs from that of B cells. We have presented here what we hope is a useful summary of current information, and hope that our speculations about the similarities and differences between TCRs and immunoglobulins will be more of a help than a hindrance.

We thank Drs Yueh-hsiu Chien, Leon Rosenberg, Leslie Berg, Kai Zinn, Peter Parham, Herman Waldmann, Michael Levitt and Don Wiley for helpful comments, Dr Ian MacNeil for help with figures and Drs Yueh-hsiu Chien, Leslie Berg, John Elliott, Hans Acha-Orbea, Barbara Fazekas de St Groth and Jane Parnes for the communication of data before publication. We also thank

Kathy Redman for preparation of the manuscript and the National Institutes of Health and the American Cancer Society for support. M.M.D. is a scholar of the PEW Foundation and P.J.B. is a fellow of the American Cancer Society.

1. Katz, D. H., Hamaoka, T. & Benacerraf, B. *J. exp. Med.* **137**, 1405-1418 (1973).
2. Rosenthal, A. S. & Shevach, E. M. *J. exp. Med.* **138**, 1194-1212 (1973).
3. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701-702 (1974).
4. Melchers, F. & Andersson, J. A. *Rev. Immunol.* **4**, 13-36 (1986).
5. Hamaoka, T. & Ono, S. A. *Rev. Immunol.* **4**, 167-204 (1986).
6. Tonegawa, S. *Nature* **302**, 575-581 (1983).
7. Eisen, H. N. *Immunology* (Harper & Row, New York, 1980).
8. Steward, M. W. & Steensgaard, J. *Antibody Affinity: Thermodynamic Aspects and Biological Significance* (CRC, Boca Raton, Florida, 1983).
9. Amzel, L. M. & Poljak, R. J. A. *Rev. Biochem.* **48**, 961-997 (1979).
10. Davies, D. R. & Metzger, H. A. *Rev. Immunol.* **1**, 87-117 (1983).
11. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. *Science* **233**, 747-753 (1986).
12. Colman, P. M. *et al. Nature* **326**, 358-363 (1987).
13. Sheriff, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 8075-8079 (1987).
14. Segal, D. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **71**, 4298-4302 (1974).
15. Amzel, L. M., Poljak, R. J., Saul, F., Varga, J. M. & Richards, F. F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1427-1431 (1974).
16. Allison, J. P., MacIntyre, B. W. & Block, D. J. *J. Immunol.* **129**, 2293-2300 (1982).
17. Haskins, K. *et al. J. exp. Med.* **157**, 1149-1169 (1983).
18. Meuer, S. C. *et al. J. exp. Med.* **157**, 705-719 (1983).
19. Samelson, L. E., Germain, R. N. & Schwartz, R. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6972-6976 (1983).
20. Kaye, J., Porcelli, S., Tite, J., Jones, B. & Janeway, Jr., C. A. *J. exp. Med.* **158**, 836-856 (1983).
21. Acuto, O. *et al. Cell* **34**, 717-726 (1983).
22. Kappler, J. *et al. Cell* **35**, 295-302 (1983).
23. Staerz, U. D., Pasternack, M. S., Klein, J. R., Benedetto, J. E. & Beban, M. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1799-1803 (1984).
24. Clevers, H., Alarcon, B., Wileman, T. & Terhorst, C. A. *Rev. Immunol.* **6**, 629-662 (1988).
25. Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. *Nature* **308**, 149-153 (1984).
26. Hedrick, S. M., Nielsen, E. A., Kvaloy, J., Cohen, D. I. & Davis, M. M. *Nature* **308**, 153-158 (1984).
27. Yang, Y. *et al. Nature* **308**, 145-149 (1984).
28. Chien, Y. *et al. Nature* **312**, 31-35 (1984).
29. Saito, H. *et al. Nature* **312**, 36-40 (1984).
30. Sim, G. K. *et al. Nature* **312**, 771-775 (1984).
31. Dembic, Z. *et al. Nature* **320**, 232-238 (1986).
32. Saito, T. *et al. Nature* **325**, 125-130 (1987).
33. Hood, L., Steinmetz, M. & Malissen, B. A. *Rev. Immunol.* **1**, 529-568 (1983).
34. Benacerraf, B. *J. Immunol.* **120**, 1809-1812 (1978).
35. Shimonkevitz, R., Kappler, J. W., Marrack, P. & Grey, H. M. *J. exp. Med.* **158**, 303-316 (1983).
36. Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. *Nature* **317**, 359-361 (1985).
37. Buus, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3968-3971 (1986).
38. Townsend, A. R. M. *et al. Cell* **44**, 959-968 (1986).
39. Guillet, J.-G., Lai, M.-Z., Briner, T. J., Smith, J. A. & Gefter, M. L. *Nature* **324**, 260-263 (1986).
40. Bjorkman, P. J. *et al. Nature* **329**, 506-512 (1987).
41. Bjorkman, P. J. *et al. Nature* **329**, 512-518 (1987).
42. Brown, J. H. *et al. Nature* **332**, 845-850.
43. Littman, D. R. A. *Rev. Immunol.* **5**, 561-584 (1987).
44. Marrack, P. *et al. J. exp. Med.* **158**, 1077-1091 (1983).
45. Gabert, J. *et al. Cell* **50**, 545-554 (1987).
46. Sleckman, B. P. *et al. Nature* **328**, 351-353 (1987).
47. Dembic, Z. *et al. Nature* **320**, 232-238 (1986).
48. Kronenberg, M., Siu, G., Hood, L. E. & Shastri, N. A. *Rev. Immunol.* **4**, 529-591 (1986).
49. Davis, M. M. *Rev. Immunol.* **3**, 537-560 (1985).
50. Saito, H. *et al. Nature* **309**, 757-762 (1984).
51. Chien, Y. *et al. Nature* **327**, 677-682 (1987).
52. Brenner, M. B. *et al. Nature* **322**, 145-149 (1986).
53. Bank, I. *et al. Nature* **322**, 179-181 (1986).
54. Pardoll, D. M. *et al. Nature* **326**, 79-81 (1987).
55. Stingl, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 2430-2434 (1987).
56. Kuziel, W. A. *et al. Nature* **328**, 263-266 (1987).
57. Hayday, A. C. *et al. Cell* **40**, 259-269 (1985).
58. Chien, Y. *et al. Nature* **330**, 722-727 (1987).
59. Patten, P. *et al. Nature* **312**, 40-46 (1984).
60. Barth, R. *et al. Nature* **316**, 517-523 (1985).
61. Arden, B., Klotz, J., Sui, G. & Hood, L. *Nature* **316**, 783-787 (1985).
62. Becker, D. M. *et al. Nature* **317**, 430-434 (1985).
63. Wu, T. T. & Kabat, E. A. *J. exp. Med.* **132**, 211-250 (1970).
64. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. *Sequences of Proteins of Immunological Interest* (NIH, Bethesda, Maryland, 1987).
65. van den Elsen, P. *et al. Nature* **312**, 413-418 (1984).
66. Gold, D. P. *et al. Nature* **321**, 431-434 (1986).
67. Krissansen, G. W., Owen, M. J., Verbi, W. & Crumpton, M. J. *EMBO J.* **5**, 1799-1808 (1986).
68. Weissman, A. M. *et al. Science* **239**, 1018-1021 (1988).
69. Raulet, D. H., Garman, R. D., Saito, H. & Tonegawa, S. *Nature* **314**, 103-107 (1985).
70. Samelson, L. E. *et al. Nature* **315**, 765-768 (1985).
71. Snodgrass, H. R., Dembic, Z., Steinmetz, M. & von Boehmer, H. *Nature* **315**, 232-233 (1985).
72. Elliott, J. R., Rock, E. P., Davis, M. M. & Chien, Y. *Nature* **331**, 627-631 (1988).
73. Akira, S., Okazaki, K. & Sakano, H. *Science* **238**, 1134-1138 (1987).
74. Yancopoulos, G. D., Blackwell, T. K., Suh, H., Hood, L. & Alt, F. W. *Cell* **44**, 251-259 (1986).
75. Fujimoto, S. & Yamagishi, H. *Nature* **327**, 242-243 (1987).
76. Okazaki, K., Davis, D. D. & Sakano, H. *Cell* **49**, 477-485 (1987).
77. Lai, E., Barth, R. K. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3846-3850 (1987).
78. Lee, N. E. & Davis, M. M. *J. Immunol.* **140**, 1665-1675 (1988).
79. Malissen, M. *et al. Nature* **319**, 28-33 (1986).
80. Iwashima, M., Green, A., Davis, M. M. & Chien, Y. *Proc. natn. Acad. Sci. U.S.A.* (in the press).



81. Alt, F. W. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4118-4122 (1982).
82. Lindsten, T., Lee, N. E. & Davis, M. M. *Proc. natn. Acad. Sci. U.S.A.* **48**, 7639-7643 (1987).
83. Chou, H. S., Nelson, C. A., Godambe, S. A., Chaplin, D. D. & Loh, D. Y. *Science* **238**, 545-548 (1987).
84. Behlke, M. A. *et al. Science* **229**, 566-570 (1985).
85. Behlke, M. A., Chou, H. S., Huppi, K. & Loh, D. Y. *Proc. natn. Acad. Sci. U.S.A.* **83**, 767-771 (1986).
86. Kotzin, B. L., Barr, V. L. & Palmer, E. *Science* **229**, 167-171 (1985).
87. Noonan, D. J. *et al. J. exp. Med.* **163**, 644-653 (1986).
88. Iwamoto, A. *et al. J. exp. Med.* **163**, 1203-1212 (1986).
89. Garman, R. D., Doherty, P. J. & Raulet, D. H. *Cell* **45**, 733-742 (1986).
90. Heilig, J. S. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8070-8074 (1987).
91. Strauss, C. *et al. Science* (1987).
92. Lindsten, T., Fowlkes, B. J., Samelson, L. E., Davis, M. M. & Chien, Y. *J. exp. Med.* **166**, 761-775 (1987).
93. Siu, G. *et al. Nature* **311**, 344-350 (1984).
94. Luria, S., Gross, G., Horowitz, M. & Givol, D. *EMBO J.* **6**, 3307-3312 (1987).
95. Quettermous, T. *et al. Nature* **322**, 184-187 (1986).
96. Winoto, A. *et al. Nature* **324**, 679-682 (1986).
97. Gorman, J. *et al. Cell* **40**, 859-867 (1985).
98. Novotny, J., Tonegawa, S., Saito, H., Kranz, D. M. & Eisen, H. N. *Proc. natn. Acad. Sci. U.S.A.* **83**, 742-746 (1986).
99. Chothia, C., Boswell, D. R. & Lesk, A. M. *Cell* (submitted).
100. Chothia, C. & Lesk, A. M. *J. molec. Biol.* **196**, 901-917 (1987).
101. Novotny, J. & Haber, E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4592-4596 (1985).
102. Chothia, C., Novotny, J., Brucoleri, R. & Karplus, M. *J. molec. Biol.* **186**, 651-663 (1985).
103. Kappler, J. W. *et al. Cell* **49**, 263-271 (1987).
104. Fink, P. J., Matis, L. A., McElligott, D. L., Bookman, M. & Hedrick, S. M. *Nature* **321**, 219-226 (1986).
105. Heber-Katz, E. *et al. J. exp. Med.* **155**, 1086-1099 (1982).
106. Mengle-Gaw, L., Conner, S., McDevitt, H. O. & Fathman, C. G. *J. exp. Med.* **160**, 1184-1194 (1984).
107. Saito, T. & Germain, R. N. *Nature* **329**, 256-259 (1987).
108. Hedrick, S. M. *et al. Science* **239**, 1541-1544 (1988).
109. Acha-Orbea, H. *et al. Cell* (in the press).
110. Matis, L. A., Cron, R. & Bluestone, J. A. *Nature* **330**, 262-264 (1987).
111. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 228-294 (1979).
112. Hood, L. E., Hunkapiller, T. & Kraig, E. in *Modern Cell Biology* (ed. MacIntosh, J. R.) 305-328 (Liss, New York, 1983).
113. Johnson, P. & Williams, A. F. *Nature* **323**, 74-76 (1986).
114. Nakauchi, H., Shinkai, Y. & Okumura, K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4210-4214 (1987).
115. Chang, C.-H. *et al. Biochemistry* **24**, 4890-4897 (1985).
116. Clark, S. J., Jefferies, W. A., Barclay, A. N., Gagnon, J. & Williams, A. F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1649-1653 (1987).
117. Goldstein, S. A. N. & Mescher, M. F. *J. Immun.* (in the press).
118. Schwartz, R. H. in *Immunogenetics* (ed. Paul, W. E.) 91-151 (Raven, New York, 1984).
119. Marrack, P. & Kappler, J. *Science* **238**, 1073-1079 (1987).
120. Jerne, N. *Eur. J. Immun.* **1**, 1-9 (1971).
121. Garman, R. D., Ko, J.-L., Vulpe, C. D. & Raulet, D. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3987-3991 (1986).
122. Suh, S. W. *et al. Proteins* **1**, 74-80 (1986).
123. Kappler, J. W., Staerz, U., White, J. & Marrack, P. C. *Nature* **332**, 35-40 (1988).
124. MacDonald, H. R. *et al. Nature* **332**, 40-45 (1988).
125. Pullen, A., Marrack, P. & Kappler, J. (in preparation).
126. Marrack, P. & Kappler, J. *Nature* **332**, 840-843 (1988).
127. Katz, M. E. & Janeway, C. A. *J. Immunol.* **134**, 2064-2070 (1985).
128. Webb, S. R. & Sprent, J. *Intern. Rev. Immunol.* **1**, 151-182 (1986).

## ARTICLES

# Star formation in giant molecular associations synchronized by a spiral density wave

S. N. Vogel\*, S. R. Kulkarni† & N. Z. Scoville†

\* Physics Department, Rensselaer Polytechnic Institute, Troy, New York 12180, USA

† Owens Valley Radio Observatory and Palomar Observatory, California Institute of Technology, Pasadena, California 91125, USA

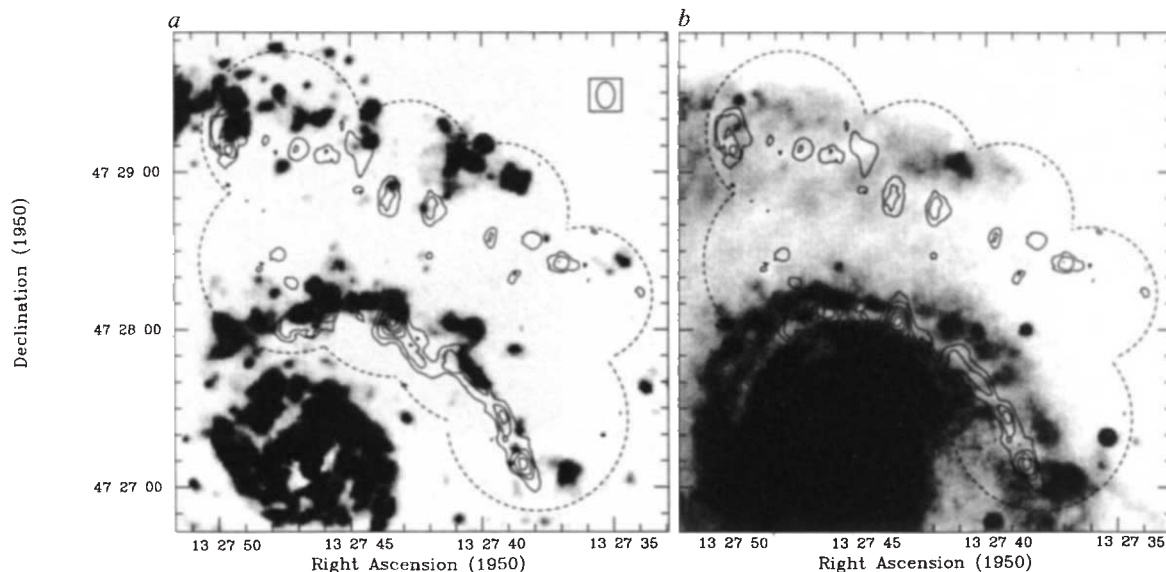
*Observations of CO emission from the prototypical spiral galaxy M51 reveal density wave streaming motions in molecular gas coincident with the dust lanes. This is strong evidence that spiral density waves assemble pre-existing molecular clouds into giant associations and trigger the collapse of suitably primed clouds, leading to the formation of stars.*

THE beautiful arms of spiral galaxies outlined by the luminous young stars have been the focus of numerous studies. In the classic description for the origin of spiral structure, the long-range gravitational potential of a stellar density wave collects and shocks atomic interstellar gas and triggers the formation of stars<sup>1,2</sup>. But molecular clouds, which may not be formed by the density wave, at least in molecule-rich galaxies, are now known to be the sites of star formation, and the evidence cited in favour of a density wave trigger can be interpreted in other ways<sup>3-5</sup>. The global star formation efficiency appears to be the same in spiral galaxies with and without grand design structure, suggesting that density waves do not trigger star formation<sup>4-6</sup>. Alternative models have emerged. For example, the arms may be produced by short-range stochastic processes, such as supernova explosions or winds from O stars which propagate star formation by compressing nearby clouds. Spiral structure then originates as a result of the shear of differential rotation in the disk, and is a transitory but constantly regenerated phenomenon<sup>3</sup>; even grand design arms such as in M51 have been reproduced by such models. Or, because the underlying stellar density waves can be much stronger than originally supposed, it may be that star formation simply appears to be enhanced as the gas clouds spend a large fraction of their time in the arms as a result of the density wave flow patterns<sup>5</sup>. Given this perplexing situation, the role of density waves in star formation is unclear.

## Molecular clouds

Although streaming motions resulting from density waves have been identified in the ionized, the atomic and the molecular gas<sup>7-9</sup>, direct evidence for a link between density wave acceleration of the sites of star formation (molecular clouds) and the subsequent birth of stars has been lacking. Such motions are most easily and unambiguously identified in external spiral galaxies, rather than in our own Galaxy. The essential requirement is for high angular resolution. Previous observations of the disks of spiral galaxies had beam sizes in the range 33-100"; much higher resolution is now available with the new mm-wave interferometers. With these issues in mind, we began a major synthesis mosaic of M51, the prototypical grand design spiral galaxy, in the 1-0 rotational transition of the CO molecule with the mm-interferometer at Owens Valley Radio Observatory (OVRO).

In Fig. 1, we present a map of the NW quadrant of M51 from  $1.5 < R < 7$  kpc obtained by combining eight 1' fields, corrected for primary beam response, and using a simple mosaic technique<sup>10</sup>. The resolution obtained was  $10 \times 7''$  (position angle  $0^\circ$ ), which corresponds to  $470 \times 330$  pc at a distance of 9.6 Mpc. These are the first observations of molecular gas in the disk of a grand design spiral galaxy with sufficient resolution to resolve the molecular cloud complexes and their streaming motions.



**Fig. 1** *a*, Contours of integrated CO (1–0) emission superimposed on a grey-scale  $H\alpha$  image of M51. The CO emission was mapped with the OVRO interferometer and is contoured at intervals of 15% of the peak of  $45 \text{ Jy km s}^{-1}$ . The area mapped in CO is outlined by the dashed line. The synthesized beam size of the CO observations is indicated in the upper right. The CO arms are seen to lie parallel to the  $H\alpha$  arms but offset toward the nucleus. This indicates that in the density wave frame newly formed massive stars lie slightly downstream from the molecular clouds out of which they formed. *b*, Contours of integrated CO emission superimposed on an R-band CCD image. The inner CO arm coincides with the well defined dust lane; the CO arm and the dust lane have a similar width of  $\sim 300 \text{ pc}$ .

In Fig. 1*a*, contours of CO emission integrated over velocity are superimposed on a 30-Å  $H\alpha$  image taken with a charge-coupled device (CCD) camera on the 60" telescope of the Palomar Observatory. The registration of the optical and radio images was done in the usual way by deriving the equatorial positions of a grid of secondary stars with respect to SAO stars, and the coordinate transformation errors are expected to be no worse than  $1''$ . In  $H\alpha$ , segments of two spiral arms can be seen spanning the interferometer field of view. A narrow CO arm, extending over more than 5 kpc, parallels the inner  $H\alpha$  arm but offset by  $\sim 7''$  towards the nucleus. There is detectable CO coinciding with the inner part of the  $H\alpha$  arm and, in this overlap region, the average velocity difference between the  $H\alpha$  emission<sup>7</sup> and CO emission is  $0 \pm 3 \text{ km s}^{-1}$ . Another less defined CO arm lies inside the outer  $H\alpha$  arm; it consists of about seven detected emission complexes spaced by about 1 kpc.

If the observed CO emission were dominated by heating effects rather than by column density, then the CO arm would coincide with the strongest heating sources—the arm of OB stars which ionize the H II regions. The offset between the  $H\alpha$  and CO arms cannot be attributed to extinction of  $H\alpha$  emission because a recent study finds that free-free emission at 6 cm, which does not suffer extinction, coincides with  $H\alpha$  emission in this region<sup>11</sup>. We conclude that the CO emission traces molecular column density in the largest complexes and that, in the frame of the density wave, the ionized peak lies downstream of the molecular peak typically by 300 pc in the inner arm and 500–1,000 pc in the outer arm.

In Fig. 1*b*, CO emission is superimposed on an R band CCD image. A narrow dust lane is apparent; it coincides with the inner CO arm to within the errors ( $1''$ ). The dust lane and the CO arm have a similar width of about 300 pc. The dust lane and CO also coincide in the outer region, although neither is as well defined here. Narrow dust lanes can be produced by either diffuse atomic or molecular clouds or dense molecular clouds compressed by a galactic density wave; simulations show that diffuse and dense molecular dust lanes need not coincide<sup>12</sup>. Our data show clearly that the dust lane coincides with the peak molecular cloud column density.

A previous interferometric study of M51, which concentrated on the nuclear region, also concluded that strong CO emission originates from the dust lanes<sup>13</sup>; in Figs 1 and 2 of ref. 13, four

CO peaks were shown coinciding with the inner dust lane mapped in our study. Three of these lie near minima of the CO arm emission mapped by us, and are probably spurious as they lie 40–80" from their northern pointing centre, well beyond the 30" half-power points of their primary beam.

### Spiral streaming motions

Perhaps the most striking new result is a systematic, resolved velocity shift across (that is, perpendicular to) the spiral arms; the shift is detected along the entire inner arm, and across the brightest complex in the outer arm. Shifts of  $-20$  to  $-30 \text{ km s}^{-1}$  across the inner arm are seen along its entire length (Fig. 2); assuming the standard inclination<sup>7</sup> of  $20^\circ$ , these correspond to velocity shifts of  $-58$  to  $-88 \text{ km s}^{-1}$  in the plane of the galaxy. In two complexes, one near the major axis and the other near the minor axis, we measured velocity gradients of  $-35$  and  $-64 \text{ km s}^{-1} \text{ kpc}^{-1}$  respectively (Fig. 3).

The velocity shifts cannot be attributed to rotation as no rotational shift is expected along the minor axis. The gradient also appears too large to be caused by a tidal warp due to the close encounter with NGC5195. On the other hand, the shifts are consistent with density wave models<sup>14,15</sup> which, as a function of position across an arm, predict an increase in the tangential velocity in the direction of rotation and a radial velocity shift toward the nucleus. The line-of-sight component of the tangential streaming velocity shift is obtained along the major axis; the observed shift is negative, which is in the sense of increasing the magnitude of the tangential velocity. The line-of-sight component of the radial velocity shift is obtained along the minor axis; it is negative, which corresponds to an inwardly directed shift (assuming as usual that the west side of M51 is the far side<sup>7</sup>). Both tangential and radial shifts are consistent with density wave theory. The streaming motions are found close to the potential minimum of the stellar density wave determined from an I-band image<sup>16</sup>, as expected.

Density wave theories predict a typical molecular surface density enhancement of factors of two to five in the arm<sup>14,15,17</sup>. The OVRO interferometer loses sensitivity to structures larger than  $\sim 25''$ ; the spiral arms, with widths under  $10''$ , are thus properly mapped. Including short-spacing data would slightly increase the arm intensities, but would not affect any of our conclusions. The CO flux contained in the CO spiral arms,



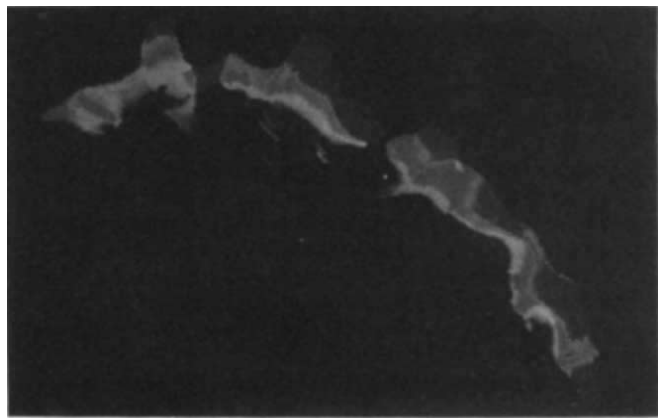
namely the dust lanes (Fig. 1), is only about 20–30% of the total flux, as measured by single-dish instruments<sup>18</sup>; this fraction is consistent with that inferred with a 33" beam<sup>9</sup> and implies that 70–80% of the molecular mass is in the interarm region, assuming a constant CO/H<sub>2</sub> conversion ratio. The peak arm enhancements in integrated CO brightness in the 330×470 pc beam are about a factor of seven above the values measured by single-dish instruments. If we subtract the arm (that is, OVRO) flux from the total (single-dish) flux, we obtain the average interarm flux, and find that the arm peaks are a factor of 10 above the average interarm flux. As is apparent in Fig. 1, the CO emission exhibits a fragmentation scale of about 1 kpc, which suggests that non-linear effects are important<sup>19–22</sup>. Averaging the interferometer fluxes over a 10" band following the arms, we derive an average arm-to-interarm integrated CO brightness ratio of 2.4 for the inner arm and 3.0 for the outer arm; the main uncertainty is the relative flux calibration of the single dish and interferometer. These enhancements appear consistent with density wave theory.

## Star formation

The mere profusion of young stars in spiral arms does not necessarily imply that density waves trigger star formation because gas flows more slowly through the arms than between the arms, which would result in an apparent enhancement in gas, and hence young stars, even without a density wave trigger<sup>2–5</sup>. There are more than 10 times as many H II regions on the downstream side of the molecular dust lane shock as on the upstream side. This enhancement is larger than the factor of 2.4–3 obtained for the molecular gas in the dust lane. For the standard models of the molecular interstellar medium, in which the majority of the mass is in giant molecular clouds (GMCs)<sup>23–25</sup>, this implies that massive star formation is enhanced more than the GMCs, which indicates that the density wave increases the rate at which those clouds form stars.

But if the molecular gas is predominantly in a diffuse state which does not form massive stars, then it is conceivable that the dust lane is produced by a hydrodynamic shock of diffuse molecular clouds, whereas the GMCs, which form the massive stars, are distributed in a broader arm coincident with H $\alpha$  emission. The GMCs would have to be enhanced by a factor of 10, like the H II regions, to obviate the need for a density wave trigger. But this picture appears unlikely for the following reasons: (1) the GMC enhancement would have to be more than three times greater than the diffuse dust lane enhancement, but CO emission outside the dust lane is observed to be three times weaker; therefore GMCs would have to contribute less than about 10% of the total CO emission to explain the weakness of CO emission. Recent observations suggest that in our Galaxy a diffuse component may contribute as much as a half to the total CO emission<sup>25</sup>; a galaxy in which the diffuse component contributed 90% of the CO flux would be unprecedented; (2) it is difficult to understand why the narrow dust lane is less enhanced than the proposed broad arm of GMCs; models show hydrodynamic arms to be narrower and more enhanced than ballistic arms<sup>2,14</sup>. If the velocity dispersions of GMCs were smaller than for diffuse clouds, this could be explained, but in our Galaxy the dispersion of diffuse clouds does not appear to be larger than for most GMCs; only the most massive GMCs have a smaller dispersion<sup>27,28</sup>; (3) the M51 dust lane appears to be slightly downstream from the stellar potential minimum inferred from the I band image; hydrodynamic calculations of strong density wave shocks invariably find the shocks upstream of the minimum, whereas the ballistic simulations find that the dust lane can be either upstream or downstream<sup>2,14</sup>. Therefore we conclude that the density wave has increased the local massive-star formation efficiency.

The streaming motions and the location and magnitude of the molecular enhancements prove that density waves exist and affect the molecular component in a manner consistent with



**Fig. 2** CO velocity structure of the inner spiral arm in M51 mapped by the OVRO interferometer, with contours and grey scale indicating velocity. Velocities calculated from the rotation curve<sup>32</sup> have been subtracted; the residual velocities are shown, with black indicating velocities greater than 13 km s<sup>-1</sup> and white indicating velocities less than -11 km s<sup>-1</sup>. The velocity is seen to shift by -20 to -30 km s<sup>-1</sup> across the arm, in a direction consistent with the streaming motions predicted by density wave theory.

theory, producing the dust lanes. We see that the well known offset<sup>29</sup> between dust lanes and H $\alpha$  is indeed an offset between star-forming molecular gas accelerated by the density wave and newly formed O stars, as to be expected if stars form a short time after the clouds have been triggered by the density wave. Furthermore, the massive-star formation efficiency is enhanced by the density wave unless CO emission arises overwhelmingly from diffuse clouds. We conclude therefore that, at least in M51, the prototypical gas-rich grand design spiral galaxy, density waves trigger massive star formation.

H I emission is offset downstream from the dust lane by about the same distance as H $\alpha$  emission, which has been interpreted as resulting from dissociation of molecular gas by recently formed stars<sup>30,31</sup>. Our data show that the narrow dust lanes mark the molecular peak and that the main H II peaks<sup>31</sup> lie downstream from the main molecular peaks, and therefore strengthen this conclusion considerably for all galaxies where an offset between the dust lanes and the star-forming H I regions is seen. We also note that the peak H I column densities<sup>31</sup> are more than an order of magnitude less than the peak molecular column densities, even when convolved to the same resolution, assuming a conversion ratio of  $N(\text{H}_2)/I(\text{CO}) = 3 \times 10^{20} \text{ cm}^{-2} \text{ K km s}^{-1}$ . Because the H I peaks lie downstream from the molecular peaks, and as the average atomic column densities are an order of magnitude lower than the molecular column densities<sup>32,33</sup>, we strongly support the claim that the classic picture of molecular clouds forming from atomic clouds compressed by a spiral density wave is incorrect, at least in molecule-rich galaxies such as M51 (ref. 13).

## Synchronizing star formation

How can the evidence presented here be reconciled with other results that indicate that density waves do not trigger star formation? The arguments against a density wave trigger are based on global comparisons: the average star formation efficiency of a galaxy does not appear to depend on the presence of a strong density wave<sup>4</sup>. Our data, on the other hand, demonstrate that density waves are an efficient local trigger: within a galaxy with a strong density wave, most of the massive-star formation is triggered by the density wave. Because only 25% of the molecular gas is in the CO lane, the global enhancement in star formation efficiency is relatively low. If we assume that locally the star formation efficiency depends on the square of the cloud density, which is consistent with our data and would be expected if cloud-cloud collisions<sup>34,35</sup> are the specific mechanism by

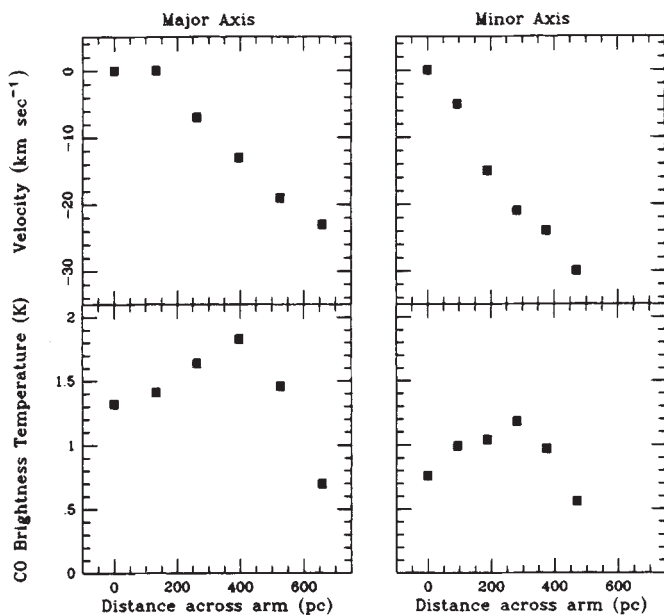


Fig. 3 CO velocities and intensities are plotted against position across a spiral arm for the CO association nearest the major axis (left) and the minor axis (right). The major axis association is at 13 h 27 min 43.2 s, 47°28'05"; the minor axis association is at 13 h 27 min 38.3 s, 47°27'10". The velocities and intensities are obtained from gaussian fits to the line profiles. The velocity scale is set to zero at the innermost arm position. Positive velocities indicate an increase in the direction of rotation and an increase away from the nucleus for the major and minor axis respectively. The observed shifts are consistent with the predictions of density wave theory.

which the density wave triggers star formation, we can estimate the global enhancement as  $S = f_{\text{arm}}\sigma_{\text{arm}}^2 + f_{\text{interarm}}\sigma_{\text{interarm}}^2$  where  $f$  is the fraction of molecular gas and  $\sigma$  is the molecular surface density. For M51,  $f_{\text{arm}} = 0.25$ ,  $f_{\text{interarm}} = 0.75$ ,  $\sigma_{\text{arm}} = 3$ , and  $\sigma_{\text{interarm}} = 1$ , which implies  $S \approx 3$ . Therefore, the average star formation efficiency is enhanced by only a factor of three compared to a galaxy without a density wave, which would be difficult to discern in the global comparisons of galaxies that have a scatter of a factor of 10 (ref. 4). Clearly, the density wave is not the only factor that affects the star formation efficiency.

But azimuthal averages of H $\alpha$  and CO emission as a function of radius in M51 show that they are linearly correlated<sup>18</sup>. For example, for  $40 < R < 120''$  (the region we observed), the total CO and H $\alpha$  emission both decrease by a factor of three. Assuming that the global star formation rate depends linearly on density, as implied by Lord's data, the opposite conclusions derived from local and global considerations can be understood if density waves do not determine which clouds will form stars, but merely trigger clouds that are already predisposed to collapse by other factors. For example, if cloud-cloud collisions are the specific mechanism by which the density wave triggers star formation, collisions would still occur in the absence of a strong density wave, but the star-formation sites would be more widely dispersed. Thus density waves merely synchronize the simultaneous collapse of many clouds, which in turn gives rise to the well defined coherent spiral arms composed of luminous young stars. A subsidiary conclusion is that the rate-limiting step in massive-star formation is not determined by the density wave: the global observations imply that the clouds whose collapse was synchronized by the density wave would ultimately form stars even without the density wave, but triggered by other, less rapid processes. The rate-limiting step—the process that primes clouds so that their collapse can be triggered by a variety of mechanisms—has yet to be identified; the global data imply the priming process is not the density wave.

Neither are the processes that occur between the triggering and the birth of young stars understood. Adopting a drift velocity

perpendicular to the arm of  $10 \text{ km s}^{-1}$  (as is reasonable from density wave models<sup>14</sup>) the observed 300 pc offset between the molecular and H $\alpha$  peaks in the inner arm translates to a delay of 30 million years. As the H $\alpha$  is produced by stars that live less than a few million years, this implies a characteristic delay of 30 million years between the density wave trigger and the appearance of massive stars. The offset is larger in the outer arm, which at first sight appears surprising as the velocity difference between the pattern and gas speeds is smaller. But the smaller velocity difference would lead to a lower shock strength, which would tend to increase the radial drift speed. Furthermore, the outer arm is probably near the co-rotation radius, where the streaming motions change qualitatively<sup>36</sup>.

## Giant molecular associations

Although three quarters of the molecular gas lies in interarm regions, our observations show unambiguously that the most massive molecular complexes lie in the dust lanes formed by the density wave, as suggested previously<sup>6</sup>. These complexes have masses in the range  $10^7$  to  $10^8 M_{\odot}$ ; because they are short-lived, we refer to them as giant molecular associations (GMAs), in analogy to OB associations. Again, assuming a  $10 \text{ km s}^{-1}$  drift velocity, from the 300 pc width of the arm we conclude that the most massive GMAs are assembled and then disrupted in  $\sim 30$  million years. We emphasize that it is not molecular clouds which are formed and destroyed, but rather the GMAs. The delay in star formation is such that most O stars do not appear until after the GMAs have been disrupted. Even though we do not detect molecular clouds coincident with most of the broad H $\alpha$  arm, they must clearly exist throughout this region as O stars cannot have drifted far from their parent clouds (much less than the 300–500 pc width of the H $\alpha$  arm). Molecular complexes could be as large as  $10^7 M_{\odot}$  and yet not be detected in our maps. Undoubtedly future efforts will concentrate on detecting these clouds, understanding the rate-limiting step, and obtaining a more accurate estimate of the observed delay intervals.

We thank the referee, B. Elmegreen, for sharpening the discussion, L. Mundy for assistance with the mosaic technique, J. Parisi for help with the data reduction, S. Lord and J. Young for providing the FCRAO CO data, J. Hester for the R and H $\alpha$  images, D. M. and B. G. Elmegreen for the I band data, S. Kleinmann and R. Tilanus for supplying positions of stars used for astrometry, and W. Roberts and F. Shu for discussion. Millimeter astronomy at OVRO is supported by the NSF. The 60" Palomar telescope is jointly operated by the California Institute of Technology and the Carnegie Institution of Washington. S.N.V. thanks the Dudley Observatory, an NSF PYI award and ATT for support.

Received 18 March; accepted 20 May 1988.

- Lin, C. C. & Shu, F. H. *Astrophys. J.* **140**, 646–655 (1964).
- Roberts, W. W. *Astrophys. J.* **158**, 123–143 (1969).
- Seiden, P. E. & Gerola, H. *Astrophys. J.* **233**, 56–66 (1979).
- Elmegreen, B. G. & Elmegreen, D. M. *Astrophys. J.* **311**, 554–562 (1986).
- Elmegreen, B. G. in *I.A.U. Symposium No. 115, Star Forming Regions* (ed. Peimbert M. & Jugaku, J.) 457–481 (Reidel, Dordrecht, 1987).
- Stark, A. A., Elmegreen, B. G. & Chance, D. *Astrophys. J.* **322**, 64–73 (1987).
- Tully, R. B. *Astrophys. J. Suppl.* **27**, 437–448 (1974).
- Visser, H. *Astr. Astrophys.* **88**, 159–174 (1980).
- Rydbeck, G., Hjalmarson, Å. & Rydbeck, O. E. H. *Astr. Astrophys.* **144**, 282–298 (1985).
- Mundy, L. G., Scoville, N. Z., Baath, L. B., Masson, C. R. & Woody, D. G. *Astrophys. J. Lett.* **304**, L51–L55 (1986).
- Tilanus, R. P. J., Allen, R. J., van der Hulst, J. M., Crane, P. C. & Kennicutt, R. C. *Astrophys. J.* (in the press).
- Elmegreen, B. G. *Astrophys. J.* **326**, 616–638 (1988).
- Lo, K. Y. et al. *Astrophys. J. Lett.* **317**, L63–L68 (1987).
- Roberts, W. W. & Hausman, M. A. *Astrophys. J.* **277**, 744–767 (1984).
- Sundelius, B., Thomasson, M., Valtonen, M. J. & Byrd, G. G. *Astr. Astrophys.* **174**, 67–77 (1987).
- Elmegreen, D. M. & Elmegreen, B. G. *Astrophys. J. Suppl.* **54**, 127–149 (1984).
- Kwan, J. & Valdes, F. *Astrophys. J.* **315**, 92–103 (1987).
- Lord, S. thesis, Univ. Massachusetts (1987).
- Mouschovias, T. Ch., Shu, F. H. & Woodward, P. *Astr. Astrophys.* **33**, 73–77 (1974).
- Balbus, S. A. & Cowie, L. L. *Astrophys. J.* **297**, 61–75 (1985).
- Balbus, S. A. *Astrophys. J.* **324**, 60–74 (1988).
- Elmegreen, B. G. *Astrophys. J.* **312**, 626–639 (1987).



23. Casoli, F., Combes, F. & Gerin, M. *Astr. Astrophys.* **133**, 99-109 (1984).
24. Sanders, D. B., Solomon, P. M. & Scoville, N. Z. *Astrophys. J.* **276**, 182-203 (1985).
25. Dame, T. M., Elmegreen, B. G., Cohen, R. S. & Thaddeus, P. *Astrophys. J.* **305**, 892-908 (1986).
26. Polk, K. S., Knapp, G. R., Stark, A. A. & Wilson, R. W. *Astrophys. J.* (in the press).
27. Stark, A. A. & Brand, J. *Astrophys. J.* (in the press).
28. Blitz, L., Magnani, L. & Mundy, L. *Astrophys. J. Lett.* **282**, L9-L12 (1984).
29. Mathewson, D. S., van der Kruit, P. C. & Brouw, W. N. *Astr. Astrophys.* **17**, 468-486 (1972).
30. Allen, R. J., Atherton, P. D. & Tilanus, R. P. J. *Nature* **319**, 296-298 (1986).
31. Tilanus, R. P. J. & Allen, R. in *Star Formation in Galaxies* (ed. Persson, C. J. L.) 309-313 (NASA, Washington, 1987).
32. Scoville, N. Z. & Young, J. S. *Astrophys. J.* **265**, 148-165 (1983).
33. Welichew, L. & Gottesman, S. T. *Astr. Astrophys.* **24**, 59-67 (1973).
34. Scoville, N. Z., Sanders, D. B. & Clemens, D. P. *Astrophys. J. Lett.* **310**, L77-L81 (1986).
35. Shu, F. H., Adams, F. C. & Lizano, S. A. *Rev. Astr. Astrophys.* **25**, 23-81 (1987).
36. Shu, F. H., Milione, V. & Roberts, W. W. *Astrophys. J.* **183**, 819-841 (1973).

# Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3

Masazumi Matsumura, Wayne J. Becktel & Brian W. Matthews

Institute of Molecular Biology, and Departments of Physics and Chemistry, University of Oregon, Eugene, Oregon 97403, USA

*Replacing the isoleucine at amino-acid position three of bacteriophage T4 lysozyme causes changes in the thermodynamic stability of the protein that are directly related to the hydrophobicity of the substituted residue. Structural analysis confirms that the hydrophobic stabilization is proportional to the reduction of the surface area accessible to solvent on folding.*

ONE of the key problems in molecular biology is to understand the physical forces that guide a polypeptide chain to its specific, folded form in an aqueous environment. Of these forces, hydrophobic interactions<sup>1-6</sup> are thought to be the most important, because other non-covalent interactions such as hydrogen bonds, van der Waals forces and electrostatic interactions are offset by similar interactions that can occur between the polypeptide and water when the protein is unfolded. Although the significance of the hydrophobic effect has been recognized and generally accepted since Kauzmann's classic paper<sup>1</sup>, the quantitative importance of hydrophobic effects at individual sites in proteins has yet to be established.

One approach to better understanding the molecular basis of protein stability is to use mutant proteins with single amino-acid substitutions<sup>7-14</sup>. Recent progress in recombinant DNA techniques, especially site-directed mutagenesis, has made it possible to modify protein structures almost at will. These techniques together with high-resolution crystal structures and measurement of thermodynamic properties should help to clarify the forces that stabilize protein structure<sup>15-21</sup>.

Here we replace the isoleucine residue at position three (Ile 3) of bacteriophage T4 lysozyme with 13 different amino-acid residues by site-directed mutagenesis. Thermodynamic measurements of stability and high-resolution X-ray structural analysis show directly how hydrophobic interactions at the Ile 3 site contribute to the overall stability of the protein.

## Mutant lysozymes

To investigate the role of Ile 3 in the stabilization of lysozyme, oligonucleotide-directed mutagenesis was used to replace Ile (ATA) of wild-type lysozyme by Trp (TGG), Phe (TTT), Tyr (TAT), Leu (TTA), Val (GTT), Met (ATG), Cys (TGT), Ala (GCT), Gly (GGT), Thr (ACA), Ser (TCT), Glu (GAA) and Asp (GAC) at position 3. Mutagenesis was carried out on the lysozyme gene cloned into phage M13mp18 using either the two-primer method of Zoller and Smith<sup>22</sup> or the uracil-template procedure of Kunkel<sup>23</sup> (for which the templates were prepared from *Escherichia coli* strain CJ236 (*dut1*, *ung1*, *thi1*, *relA1*/pCJ105(Cm<sup>r</sup>) kindly provided by Dr Kunkel). The desired mutations were identified by DNA sequencing<sup>24</sup> and the mutant genes were subcloned into the expression vector pHSe5, which contains the *tac* and *lacUV5* promoters<sup>25</sup>. Expression was induced by the addition of isopropyl- $\beta$ -thiogalactoside and the mutant proteins were purified<sup>25,26</sup> and shown to be homogeneous

by SDS-gel electrophoresis<sup>27</sup> and reverse-phase high-pressure liquid chromatography.

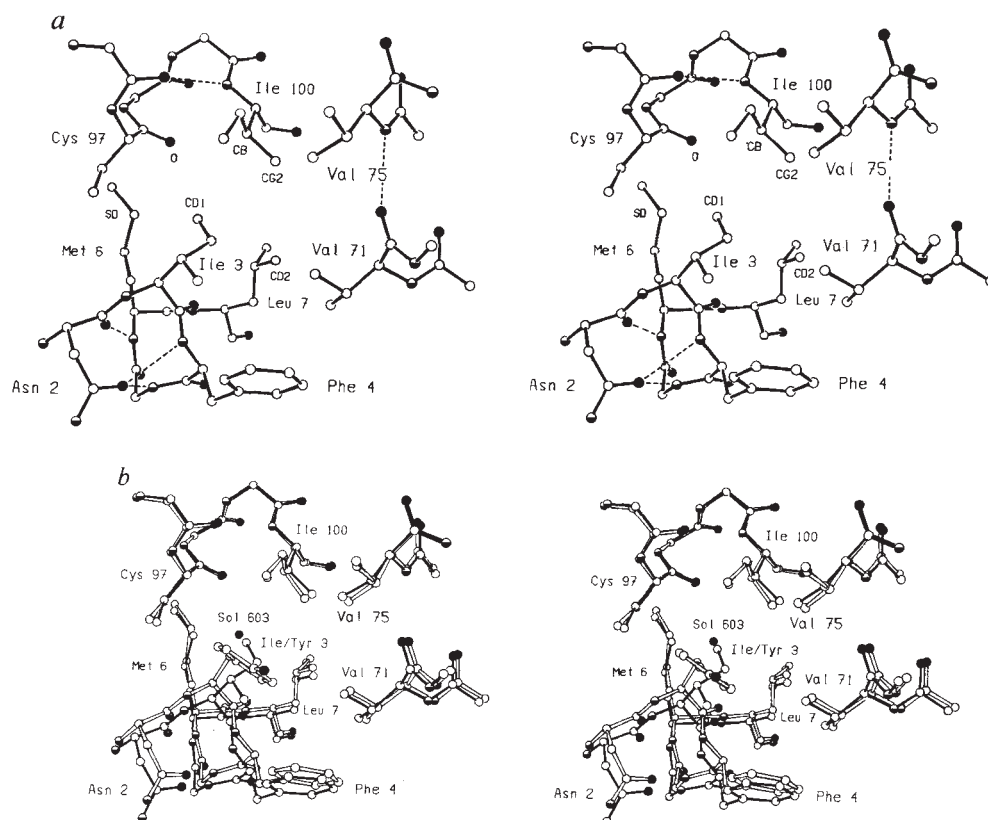
Attempts were made to crystallize all of the mutant proteins with phosphate as the precipitant, as used for the wild-type lysozyme<sup>28,29</sup>, but this was successful for only two of the mutant lysozymes, Ile 3  $\rightarrow$  Val (I3V) and Ile 3  $\rightarrow$  Tyr (I3Y). X-ray diffraction data to 1.7-Å resolution were collected for these crystals by oscillation photography and electron-density maps with coefficients ( $F_{\text{mutant}} - F_{\text{wild-type}}$ ) and ( $2F_{\text{mutant}} - F_{\text{wild-type}}$ ) and phases of the refined wild-type structure<sup>29</sup> were used to build an initial model of the mutant structure on a PS300 graphics system, using the program FRODO (ref. 30). The models of each mutant structure were refined to crystallographic *R*-values of ~15.7% using the TNT package of programs<sup>31</sup>. (Details of the crystallographic analyses will be given elsewhere.)

In wild-type lysozyme, Ile 3 contributes to the major hydrophobic core of the C-terminal lobe and also helps link the N- and C-terminal domains (for example, see Fig. 3 of ref. 28). The side chain of residue 3 is ~80% inaccessible to solvent and contacts the side chains of Met 6, Leu 7 and Ile 100 as well as the main chain of Cys 97 (Fig. 1a). These four residues are buried within the protein interior. The structure of the mutant lysozyme with Ile 3 replaced by Val is very similar except for the deletion of the  $\sigma$ -methyl group. In contrast, the replacement of Ile 3 with tyrosine is associated with substantial changes in structure (Fig. 1b). Perhaps the most striking feature is that the side chain of the tyrosine is not accommodated within the interior of the protein but rotates ~120° so that it is largely exposed to the solvent. Removal of the side chain of Ile 3 tends to create a cavity adjacent to the hydrophobic core. In the Tyr 3 mutant protein this cavity is occupied in part by a bound water molecule. Also, the  $\alpha$ -,  $\beta$ - and  $\gamma$ -carbon atoms of Tyr 3 move 0.7-1.7 Å into the hydrophobic cavity (relative to their counterparts in the wild-type (Ile 3) structure) (Fig. 1b). In apparent association with this movement, the N-terminal polypeptide (residues 1-9) shifts from 0.6 Å to 1.1 Å. As will be discussed elsewhere, the shift in the N-terminal residues provides a simple explanation for the finding that crystals were obtained only for the variants with isoleucine (wild-type), valine and tyrosine at position 3.

## Thermal stability and hydrophobicity

Bacteriophage T4 lysozyme can be reversibly unfolded (denatured) by heating both at acidic and neutral pH<sup>10,32,33</sup>. At

**Fig. 1** *a*, Stereo drawing showing the environment of Ile 3 in wild-type lysozyme. Residues 3–7 are within one  $\alpha$ -helix, residues 97–100 lie in another  $\alpha$ -helix and residues 71–75 form part of a third  $\alpha$ -helix. Those atoms that are within 4 Å of the side chain of Ile 3 are labelled individually. Carbon atoms are shown as open circles, nitrogen atoms are half filled and oxygen atoms are drawn solid. *b*, Superposition of the structure of the mutant lysozyme Ile 3  $\rightarrow$  Tyr (open bonds) on that of wild-type lysozyme (solid bonds). The solvent molecule (number 603) is present only in the mutant structure. Before making this superposition the coordinates of the mutant molecule were rotated and translated to minimize the overall difference between the mutant and wild-type structures.



$T_m$ , the temperature of the midpoint of the thermal denaturation transition, half the protein is unfolded, and the apparent free energy of unfolding ( $\Delta G$ ) is zero. The changes in  $T_m$  both at pH 2.0 and at pH 6.5, caused by mutations at position 3, are shown in Table 1. The corresponding changes in the free energy of stabilization due to the different mutations are also given.

Because the side chain of Ile 3 contributes to the major hydrophobic core of the lysozyme molecule it seems likely that the hydrophobicity of the residue at position 3 contributes to the stability of the molecule. A number of hydrophobicity scales of amino-acid residues have been proposed<sup>34–41</sup>, however, and there is some controversy as to the best way in which to quantify this effect. Following earlier studies<sup>40,42,43</sup>, Rose *et al.*<sup>4</sup> showed that there is a strong correlation between the tendency for residues to be buried on folding, and hydrophobicity scales assessed in a water-ethanol transfer system<sup>34,35</sup>. As a first trial, therefore, we used this classic scale to compare the hydrophobicity of the residue at position 3 with protein stability (Table 1, Fig. 2a). There is a clear correlation between residue hydrophobicity ( $\Delta G_{tr}$ , kcal mol<sup>-1</sup>) and protein stability at pH 2.0 ( $\Delta \Delta G$ , kcal mol<sup>-1</sup>). Except for mutant proteins with Phe, Tyr, Trp and Cys (S-S), the data on 11 proteins display an essentially linear relationship. The reason why the Tyr mutant, and presumably also Phe and Trp, diverge from linearity is immediately apparent from the observed structure of I3Y (Fig. 1b). The side chain of the tyrosine protrudes into the solvent and remains partially solvated in the folded form, and thus does not exhibit its full hydrophobic potential. As described by Perry and Wetzel<sup>45</sup> a cysteine at position 3 of T4 lysozyme forms a disulphide linkage with Cys 97 to give a more stable protein than the wild type. The enhanced stability of this variant is, presumably, due at least in part to entropic effects.

The correlation between residue hydrophobicity and protein stability seen in Fig. 2a means that each unit increase in the side chain hydrophobicity at position 3 is associated with a constant increase in the overall stability of the protein molecule. In other words, the data suggest that the local hydrophobic interactions at position 3 directly contribute to the overall ther-

mal stability of T4 lysozyme. Furthermore, the slope of the straight line in Fig. 2a is about 0.8, indicating that 1 kcal mol<sup>-1</sup> of water/ethanol transfer free energy contributes  $\sim 0.8$  kcal mol<sup>-1</sup>, on average, to the overall stability of the protein. This shows that hydrophobic interaction is the major factor governing protein stabilization at position 3.

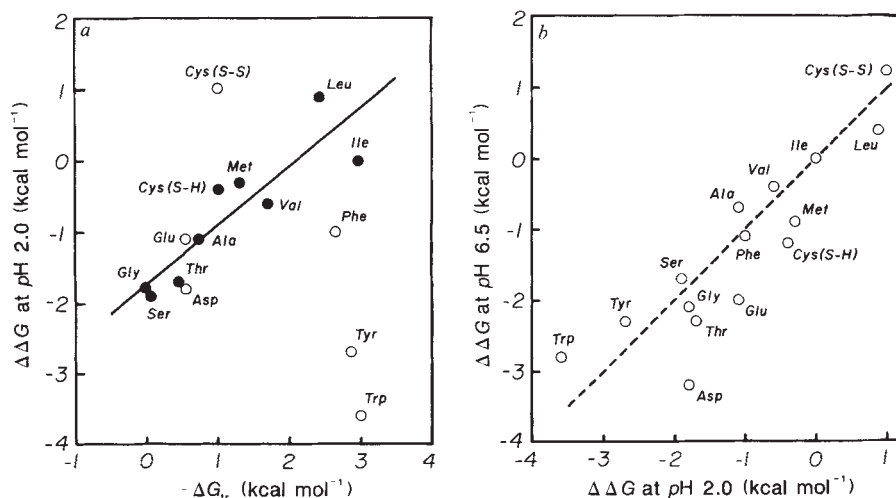
**Table 1** Thermal stability of lysozymes with different amino acids at position 3

| Amino acid at position 3 | $\Delta T_m$ (°C) |        | $\Delta \Delta G$ (kcal mol <sup>-1</sup> ) |        | $-\Delta G_{tr}$ (kcal mol <sup>-1</sup> ) |
|--------------------------|-------------------|--------|---------------------------------------------|--------|--------------------------------------------|
|                          | pH 2.0            | pH 6.5 | pH 2.0                                      | pH 6.5 |                                            |
| Trp                      | -16.4             | -8.0   | -3.6                                        | -2.8   | 3.00                                       |
| Ile                      | 0.0               | 0.0    | -0.0                                        | 0.0    | 2.97                                       |
| Tyr                      | -9.5              | -5.9   | -2.7                                        | -2.3   | 2.87                                       |
| Phe                      | -4.0              | -3.0   | -1.0                                        | -1.1   | 2.65                                       |
| Leu                      | 3.0               | 0.9    | 0.9                                         | 0.4    | 2.42                                       |
| Val                      | -2.1              | -1.2   | -0.6                                        | -0.4   | 1.69                                       |
| Met                      | -0.9              | -2.3   | -0.3                                        | -0.9   | 1.30                                       |
| Cys (S-H)                | -1.9              | -3.7   | -0.4                                        | -1.2   | 1.00                                       |
| Cys (S-S)                | 4.8               | 3.3    | 1.0                                         | 1.2    | 1.00                                       |
| Ala                      | -3.8              | -1.8   | -1.1                                        | -0.7   | 0.73                                       |
| Thr                      | -6.1              | -6.0   | -1.7                                        | -2.3   | 0.44                                       |
| Ser                      | -7.0              | -4.6   | -1.9                                        | -1.7   | 0.04                                       |
| Gly                      | -7.2              | -5.8   | -1.8                                        | -2.1   | 0.00                                       |
| Glu                      | -4.1              | -5.7   | -1.1                                        | -2.0   | 0.55 (-1.63)                               |
| Asp                      | -6.5              | -8.5   | -1.8                                        | -3.2   | 0.54 (-2.63)                               |

$\Delta T_m$  is the change in melting temperature of the mutant lysozyme ( $\pm 0.5^\circ$ ) relative to wild-type lysozyme ( $T_m = 41.9^\circ\text{C}$  at pH 2.0;  $T_m = 64.7^\circ\text{C}$  at pH 6.5).  $\Delta \Delta G$  is the difference between the free energy of unfolding of the mutant protein and that of wild type at the melting temperature of wild-type lysozyme. A positive value of  $\Delta \Delta G$  indicates that the mutant is more stable than wild type.  $-\Delta G_{tr}$  is the free energy of transfer from water to ethanol. Values for  $\Delta G_{tr}$  are taken from ref. 34 except for Cys (S-S) which is from ref. 41. The values in parentheses correspond to the ionized forms of the side chains and were calculated by the method of Levitt<sup>44</sup>.



**Fig. 2** *a*, Free energy of stabilization ( $\Delta\Delta G$ ) of mutant lysozymes at pH 2.0, plotted against the free energy of transfer ( $-\Delta G_{tr}$ ) of individual amino acids from water to ethanol (from ref. 34). Protein free energies are plotted relative to wild type (Ile) and transfer free energies are relative to glycine. The straight line was determined by regression analysis using the nine amino acids indicated by solid circles. Amino acids omitted from the analysis are the acidic residues, Glu and Asp, the large aromatic amino acids Phe, Tyr and Trp and Cys (S-S) (see text). *b*, Free energy of stabilization of mutant lysozymes at pH 6.5 plotted against stabilization at pH 2.0.



One of the characteristics of the hydrophobic interaction is that it is independent of pH. Figure 2*b* shows protein stability ( $\Delta\Delta G$ ) at pH 2.0 plotted against that at pH 6.5. As can be seen, most of the values approximate a straight-line relationship of unit slope. This indicates that the changes in the protein stabilities at pH 2.0 are approximately equivalent to those at pH 6.5. In other words, the change in protein stability associated with mutations, as measured by the change in  $\Delta G$ , is independent of pH. Two of the most severe departures from linearity occur for the mutant proteins substituted with Asp and Glu (Fig. 2*b*). Because the  $pK_a$  values of the side chains of Asp and Glu are presumably in the range 4.4–4.6 (ref. 41), it is likely that the carboxyl groups of these residues are, more or less, ionized at pH 6.5 but not at pH 2.0. It is known that the charged forms of amino-acid side chains have enhanced solubilities in water, and therefore decreased hydrophobicities<sup>41</sup>. The hydrophobicities of the charged forms of Asp and Glu at pH 6.5 were therefore calculated according to the method of Levitt<sup>44</sup>, and are given in parentheses in Table 1. The loss of protein stability at pH 6.5 relative to pH 2.0 for the Glu and especially for the Asp replacements are consistent with the decreased hydrophobicities of these residues.

It is of interest to compare the observed stability of the mutant proteins with other hydrophobicity scales. Fauchere and Plaska<sup>36</sup> determined the hydrophobic parameter of amino-acid side chains from partitioning of N-acetyl-amino-acid amide derivatives between water and octanol. Damodaran and Song<sup>37</sup> determined the free energies of transfer of amino acids from water to N-methylacetamide, which has a higher dielectric constant than water. Solubility measurements were used by Feudler *et al.*<sup>38</sup> to assess free energies of transfer of amino acids from water to hexane. Bull and Breese<sup>39</sup> measured the surface tension of amino acids and calculated the transfer free energies from the bulk to the surface of an aqueous solution. In contrast to the water-solvent models mentioned above, Wolfenden *et al.*<sup>40</sup>

determined transfer free energies of amino-acid side chains between aqueous solutions and the vapour phase.

As shown in Table 2, protein stability is highly correlated with each of these proposed hydrophobicity scales except that from the water-vapour model<sup>40</sup>. The discrepancy in the latter case may occur because water-vapour partition free energies can include contributions from hydrogen bonding and dispersion forces, in addition to hydrophobic effects. The consistency of the scales from different water-solvent model systems agrees with the view that free energy of transfer is not very sensitive to the choice of organic solvent because the principal factor being measured in such transfers is the energetic cost of removing solute from water<sup>34,35</sup>. Another feature of this analysis is that the slope of the correlation line is fairly constant (0.81–0.94), except for the hexane and vapour systems. Allowing for the fact that Ile 3 has about 80% of its side chain surface area inaccessible to solvent, it appears that the free energy contribution of a fully-buried hydrophobic residue to protein stability is approximated very well by the free energy of transfer from water to organic solvent. Yutani *et al.*<sup>13</sup> also found a linear correlation between hydrophobicity and stability for a series of variants of tryptophan synthase, although in this case the slope was substantially greater than unity and varied with pH, suggesting that factors other than hydrophobicity were involved.

### Surface area and protein stability

Using the concept of accessible surface area, as introduced by Lee and Richards<sup>46</sup>, Chothia correlated the surface area of amino-acid residues with the energy of transfer from water to organic solvent<sup>47</sup>. The results suggested that the changes in the surface area that occur during protein folding can be used to calculate the contribution of hydrophobic interactions to the free energy of folding.

This concept can be applied to calculate the free energy of the hydrophobic interaction of the mutant lysozymes, but a

**Table 2** Correlations of lysozyme stability with various hydrophobicity scales

| Transfer process | Method          | Number of residues examined | Correlation coefficient | Slope | Reference                             |
|------------------|-----------------|-----------------------------|-------------------------|-------|---------------------------------------|
| Water to ethanol | Solubilities    | 9                           | 0.89                    | 0.81  | Tanford <sup>34</sup>                 |
| Water to octanol | Solubilities    | 9                           | 0.94                    | 0.91  | Fauchere and Pliska <sup>36</sup>     |
| Water to NMA*    | Solubilities    | 9                           | 0.92                    | 0.94  | Damodaran and Song <sup>37</sup>      |
| Water to hexane  | Distributions   | 7†                          | 0.92                    | 0.65  | Feudler <i>et al.</i> <sup>38</sup>   |
| Water to surface | Surface tension | 9                           | 0.86                    | 0.88  | Bull and Breese <sup>39</sup>         |
| Water to vapour  | Vapour pressure | 9                           | 0.51                    | 0.16  | Wolfenden <i>et al.</i> <sup>40</sup> |

Changes in stability of lysozymes ( $\Delta\Delta G$  at pH 2.0, in Table 1) for Ile, Leu, Val, Met, Cys (S-H), Ala, Thr, Ser and Gly were used for the calculation. Correlation coefficient and slope were obtained from a least-squares linear regression.

\* N-methylacetamide.

† Data for Met and Cys (S-H) are not included.

**Table 3** Calculation of the hydrophobic free energy of stabilization of lysozyme due to the transfer of residue 3 from a fully solvated form to the interior of the folded protein

| Amino acid residue at position 3 | $A_0$<br>( $\text{\AA}^2$ ) | $A$<br>( $\text{\AA}^2$ ) | $A_0 - A$<br>( $\text{\AA}^2$ ) | $A_0 - A$ (Gly)<br>( $\text{\AA}^2$ ) | $\Delta\Delta G_{\text{calc}}$<br>(kcal mol $^{-1}$ ) | $\Delta\Delta G_{\text{obs}}$<br>(kcal mol $^{-1}$ ) |
|----------------------------------|-----------------------------|---------------------------|---------------------------------|---------------------------------------|-------------------------------------------------------|------------------------------------------------------|
| Ile                              | 151.6                       | 20.5                      | 131.1                           | 91.4                                  | 2.2                                                   | 1.8–2.1                                              |
| Val                              | 128.7                       | 17.6                      | 111.1                           | 71.4                                  | 1.7                                                   | 1.2–1.7                                              |
| Ala                              | 87.1                        | 14.8                      | 72.3                            | 32.6                                  | 0.8                                                   | 0.7–1.4                                              |
| Gly                              | 56.3                        | 16.6                      | 39.7                            | 0.0                                   | 0.0                                                   | 0.0                                                  |

Solvent-accessible surface area was calculated by the method of Lee and Richards<sup>46</sup> using a rolling water molecule of radius 1.4 Å.  $A_0$  is the solvent-accessible surface area of the residue in an extended ( $\phi = -139^\circ$ ,  $\psi = 135^\circ$ ) tripeptide Asn-X-Phe.  $A$  is the solvent-accessible surface area of residue 3 in the folded protein. The difference in solvent exposure ( $A_0 - A$ ) between extended ( $A_0$ ) and folded ( $A$ ) state is assumed to correspond to the area that is buried when lysozyme folds.  $A_0 - A$  (Gly) is the value of ( $A_0 - A$ ) after subtraction of the accessible surface area of glycine (39.7 Å<sup>2</sup>).  $\Delta\Delta G_{\text{calc}}$ , the calculated free energy of folding relative to glycine, is obtained by multiplying the previous column by 24 cal Å<sup>-2</sup> mol<sup>-1</sup> (ref. 47). The observed free energy of folding ( $\Delta\Delta G_{\text{obs}}$ ), relative to glycine, is from Table 1. Coordinates for the calculation of the solvent-accessible area  $A$  were from crystal structures in the case of Ile and Val and were inferred from model building in the other cases (see text).

standard state is needed to do this. The reference state of an amino-acid residue X is often approximated by an extended tripeptide of the form Gly-X-Gly or Ala-X-Ala<sup>4,42,46</sup>. In this case we used the same procedure except that the sequence Asn-X-Phe was substituted to correspond to residues 2–4 of T4 lysozyme (compare with ref. 48). The surface area accessible to solvent ( $A_0$ ) in this hypothetical reference (unfolded) state was calculated<sup>46</sup> for isoleucine and other hydrophobic residues (Table 3). Table 3 also lists the corresponding values for Ile and Val in the folded form, calculated using the coordinates of the respective X-ray crystal structures.

As shown in Table 3, Ile 3 buries 131.1 Å<sup>2</sup> of surface area in going from the extended form to the folded structure. The corresponding value for Val 3 is 111.1 Å<sup>2</sup>, which is 20 Å<sup>2</sup> smaller than that for Ile. Chothia<sup>47</sup> estimated the proportionality constant between the solvent-exposed surface area of amino-acid residues and hydrophobic free energy as 24 cal Å<sup>-2</sup> mol<sup>-1</sup>. On this scale, the difference of 20 Å<sup>2</sup> between Ile and Val corresponds to 0.48 kcal mol<sup>-1</sup> of free-energy change of stabilization due to the hydrophobic effect. This calculated energy agrees surprisingly well with the observed difference of stability between wild-type (Ile 3) and mutant (Val 3) protein, that is 0.6 kcal mol<sup>-1</sup> at pH 2 and 0.4 kcal mol<sup>-1</sup> at pH 6.5.

To explore further the relation between the solvent-accessible surface area and the free energy of stabilization, hypothetical structures of the mutant lysozymes with Ala and Gly at position 3 were obtained by deleting the appropriate atoms from the wild-type (Ile 3) structure. Using the same procedure mentioned above, the solvent-accessible areas of these 'mutant' proteins with Ala and Gly were calculated for the folded and unfolded forms. The results are also shown in Table 3. Here again, the calculated free energies of hydrophobic stabilization are in remarkably good agreement with the observed protein stabilities. The attempt to infer the Ala and Gly structures from that of wild-type lysozyme is, of course, open to criticism. It can be noted, however, that the structures of many other mutant lysozymes have been shown to be very similar to wild type<sup>7–9</sup>. Also, it makes practically no difference to the calculated solvent-exposed surface area if the Ala and Gly structures are inferred from the Tyr 3 variant rather than the wild type structure. (Solvent exposure, calculated with a 'rolling water molecule' of radius 1.4 Å is relatively insensitive to localized changes in packing.) If we correlate over the areas and the free energies for the variants listed in Table 3 we obtain a proportionality constant of  $20 \pm 2$  cal Å<sup>-2</sup> mol<sup>-1</sup>. Richards<sup>48</sup> also suggested a value of 20 cal Å<sup>-2</sup> mol<sup>-1</sup> on the basis that the interior of a protein was intermediate between liquid hydrocarbon and an amino-acid crystal. It has to be remembered, however, that the proportionality constant will depend on the nature of the aqueous and the non-aqueous phases between which transfer occurs. In the case of residue 3 of lysozyme, the side-chain environment in the folded protein is mostly, although not

entirely, hydrophobic (Fig. 1a). In such a case, simple calculations based on solvent accessibility, using a proportionality constant of 20 cal Å<sup>-2</sup> mol<sup>-1</sup> for hydrophobic residues, appear to provide a realistic estimate of the hydrophobic energy of protein stabilization.

It should be noted that the replacement of an amino acid in a protein can change the solvent-accessible surface area not only of the substituted residue itself but also of other residues, especially those that are adjacent in the folded structure. We therefore extended the calculations shown in Table 3 to estimate the total change in solvent-accessible surface area of residue 3 plus the nine surrounding residues. The results were very similar to those given in Table 3 (details will be given elsewhere).

We do not mean to suggest that the free energy changes associated with substitutions of amino-acid residues can, in all cases, be determined by simple calculations of solvent accessibility, as used here. Ile 3 of T4 lysozyme is largely inaccessible to solvent, but, at the same time, is sufficiently close to the surface that different substitutions can be accommodated without requiring major changes in the protein structure, at least as judged by the crystal structure of the Val and Tyr variants. The location of Ile 3 close to the amino terminus may contribute to this adaptability. In cases where an amino-acid substitution introduces unfavourable steric interactions, or results in unsatisfied hydrogen bonds or uncompensated charges or differences in van der Waals interactions, or causes substantial changes in the conformation of the protein, other energy terms, in addition to the hydrophobic effect, must be taken into account. It should be remembered that the model used for the unfolded state assumes that the denatured protein, or at least the residue substituted, is maximally accessible to solvent. This would not necessarily be applicable if the unfolded protein retained some local structure.

Previous studies of the hydrophobic effect in proteins of known three-dimensional structure have correlated amino-acid hydrophobicity with the average area of an amino acid that is buried<sup>4</sup> and with the distribution of different types of amino acids between the protein surface and interior<sup>42,43,49</sup>. These studies have given useful insights into the role of an 'average' amino acid of a given type, but provide little information on individual residues. The present analysis, in contrast, was designed to show how the local hydrophobic effect at a specific site contributes to the overall stability of the protein.

It should be noted that the results obtained here are supportive of a number of recent conclusions concerning the effects of amino-acid substitutions on protein stability<sup>50</sup>. In the present instance 13 variants were constructed at position 3. In most cases the stabilities of the variants were somewhat less than the wild-type protein but, in each case, a folded, functional protein was obtained, consistent with the notion that proteins are remarkably tolerant of amino-acid substitutions<sup>9,20,21</sup>. Two of the variants, Ile 3 → Leu and Ile → 3 Cys (S-S) are more thermo-



stable than wild type, showing that the protein has not been optimized for maximal thermostability. The enhanced stability of the Leu variant appears to be due to an increase in hydrophobic stabilization, relative to wild type, as will be discussed elsewhere. Another of the variants, Ile 3  $\rightarrow$  Trp, is one of the least stable lysozymes characterized to date<sup>26,29</sup>. The side chain of Ile 3 has relatively low crystallographic thermal factors (see Fig. 1a of ref. 51). The present demonstration that changes in this residue can substantially alter the thermostability of the protein is consistent with the view that the rigid parts of folded proteins make the largest contribution to stability<sup>51,52</sup>. It is

particularly encouraging to find that the relative stabilities of variant proteins with different hydrophobic residues can be estimated by a straightforward calculation based on the difference between the solvent-exposed surface area in the three-dimensional structure of the variant protein and in a hypothetical unfolded structure.

We thank Joan Wozniak for technical assistance and Sun Dao-pin, Rick Faber and Larry Weaver for help with the crystallographic aspects of the analysis. This work was supported in part by grants from the NIH and the NSF (B.W.M., J. A. Schellman) and the Lucille P. Markey Charitable Trust.

Received 27 April; accepted 23 May 1988.

- Kauzmann, W. *Adv. Protein Chem.* **14**, 1-63 (1959).
- Privalov, P. L. *Adv. Protein Chem.* **33**, 167-241 (1979).
- Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
- Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. *Science* **229**, 834-838 (1985).
- Eisenberg, D. & McLachlan, A. D. *Nature* **319**, 199-203 (1986).
- Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8069-8072 (1986).
- Grütter, M. G., Hawkes, R. B. & Matthews, B. W. *Nature* **277**, 667-669 (1979).
- Grütter, M. G., Weaver, L. H., Gray, T. M. & Matthews, B. W. in *Bacteriophage T4* (eds Matthews, C. K., Kutter, E. M., Mosig, G. & Berget, P. M.) 356-360 (Am. Soc. Microbiology, Washington D.C., 1983).
- Alber, T. *et al. UCLA Symp. molec. cell. Biol. New Ser.* **39**, 307-318 (1986).
- Hawkes, R., Grütter, M. G. & Schellman, J. J. *J. molec. Biol.* **175**, 195-212 (1984).
- Fermi, G. & Perutz, M. F. in *Haemoglobin and Myoglobin* 97-101 (Oxford Univ. Press, Oxford, 1981).
- Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5685-5789 (1984).
- Yatani, K., Ogasahara, K., Tsujita, T. & Sugino, Y. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4441-4444 (1987).
- Matsumura, M., Yasumura, S. & Aiba, S. *Nature* **323**, 356-358 (1986).
- Wells, J. A. & Powers, D. B. *J. biol. Chem.* **261**, 6564-6570 (1986).
- Katz, B. A. & Kossiakoff, A. J. *J. biol. Chem.* **261**, 15480-15485 (1986).
- Bryan, P. N. *et al. Proteins: Str. Funct. Genet.* **1**, 326-334 (1986).
- Villafranca, J. E., Howell, E. E., Oatley, S. J., Xuong, N. & Kraut, J. *Biochemistry* **26**, 2182-2189 (1987).
- Matthews, B. W., Nicholson, H. & Becktel, W. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6663-6667 (1987).
- Alber, T. *et al. Nature* **330**, 41-46 (1987).
- Alber, T. *et al. Science* **239**, 631-635 (1988).
- Zoller, M. J. & Smith, M. *Meth. Enzym.* **100**, 468-500 (1983).
- Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488-492 (1985).

- Sanger, F., Nickelsen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. *Meth. Enzym.* (in the press).
- Alber, T. & Matthews, B. W. *Meth. Enzym.* **154**, 511-533 (1987).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Remington, S. J. *et al. J. molec. Biol.* **118**, 81-98 (1978).
- Weaver, L. H. & Matthews, B. W. *J. molec. Biol.* **193**, 189-199 (1987).
- Jones, T. A. in *Crystallographic Computing* (ed. Sayre, D.) 303-317 (Oxford Univ. Press, Oxford, 1982).
- Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. *Acta Crystallogr. A* **43**, 489-503 (1987).
- Elwell, M. & Schellman, J. A. *Biochem. Biophys. Acta* **386**, 309-323 (1975).
- Becktel, W. J. & Baase, W. A. *Biopolymers* **26**, 619-623 (1987).
- Tanford, C. J. *Am. chem. Soc.* **84**, 4240-4247 (1962).
- Nozaki, Y. & Tanford, C. J. *J. biol. Chem.* **246**, 2211-2217 (1971).
- Fauchere, J.-L. & Pliska, V. *Eur. J. med. Chem.-Chim. Ther.* **18**, 369-375 (1983).
- Damodaran, S. & Song, K. B. *J. biol. Chem.* **261**, 7220-7222 (1986).
- Feudler, J. H., Nome, F. & Nagyvary, J. J. *J. molec. Evol.* **6**, 215-232 (1975).
- Bull, H. B. & Breese, K. *Arch. Biochem. Biophys.* **161**, 655-670 (1974).
- Wolfenden, R., Andersson, L., Cullis, P. M. & Southgate, C. C. B. *Biochemistry* **20**, 849-855 (1981).
- Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry* (Freeman, New York, 1980).
- Chothia, C. J. *J. molec. Biol.* **105**, 1-14 (1976).
- Janin, J. *Nature* **277**, 491-492 (1979).
- Levitt, M. J. *J. molec. Biol.* **104**, 59-107 (1976).
- Perry, L. J. & Wetzel, R. *Science* **226**, 555-557 (1984).
- Lee, B. & Richards, F. M. *J. molec. Biol.* **55**, 379-400 (1971).
- Chothia, C. *Nature* **248**, 338-339 (1974).
- Richards, F. M. A. *Rev. Biophys. Bioeng.* **6**, 151-176 (1977).
- Muller, S., Janin, J., Lesk, A. M. & Chothia, C. J. *J. molec. Biol.* **196**, 641-656 (1987).
- Matthews, B. W. *Biochemistry* **26**, 6885-6888 (1987).
- Alber, T., Dao-pin, S., Nye, J. A., Muchmore, D. C. & Matthews, B. W. *Biochemistry* **26**, 3754-3758 (1987).
- Creighton, T. E. *Biopolymers* **22**, 49-58 (1983).

## LETTERS TO NATURE

### The inner radio jet of 3C273

J. A. Zensus\*, L. B. Bååth†\*, M. H. Cohen\* & G. D. Nicolson‡

\* Owens Valley Radio Observatory, California Institute of Technology, 105-24, Pasadena, California 91125, USA

† Onsala Space Observatory, 43900 Onsala, Sweden

‡ Hartebeesthoek Radio Observatory, PO Box 3718, 2000 Johannesburg, South Africa

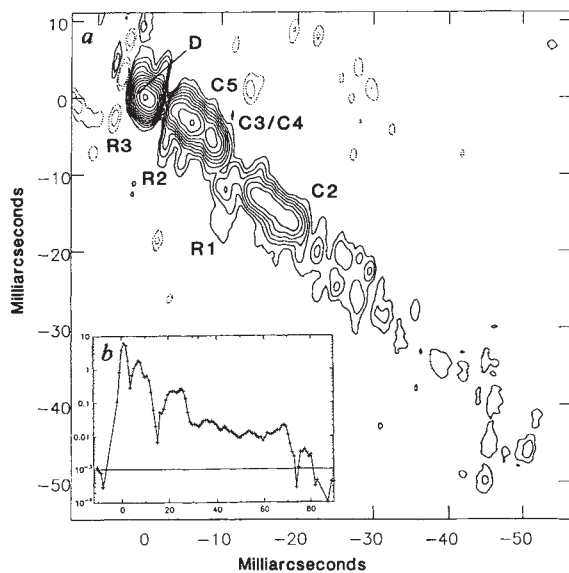
Radio maps of 3C273 obtained with very long baseline interferometry (VLBI) have been limited by low dynamic range and poor north-south resolution resulting from the low declination ( $2^\circ$ ) of this quasar<sup>1</sup>. Dramatic improvement can now be achieved using larger arrays and antennas in the Southern Hemisphere<sup>2,3</sup>. A new VLBI map, made at 5 GHz with angular resolution and dynamic range unsurpassed at this frequency for this source, shows a narrow jet extending to a projected distance  $l_{\text{proj}} \sim 125 h^{-1}$  parsecs from the core. Superluminal motion exists out to at least  $l_{\text{proj}} \approx 46 h^{-1}$  parsecs. Successive superluminal components emerge from the core and appear to move on a fixed curved path with similar speeds of about 1 milliarcseconds per year.

Figure 1a shows a map of 3C273 at 5 GHz, made from a 14-station, full ( $u, v$ )-track observation at epoch 1985.41. Technical details of this experiment will be described elsewhere. This

map is the best VLBI image of 3C273 yet made at 5 GHz. The north-south resolution and dynamic range are superior when compared with earlier maps because two southern antennas, Hartebeesthoek (South Africa) and Arecibo (Puerto Rico), were used.

Our map shows a narrow jet<sup>4</sup> extending over  $\sim 70$  mas, almost three times the length observed before at this frequency<sup>1</sup>; this corresponds to a projected scale  $l_{\text{proj}} \sim 125 h^{-1}$  pc ( $z = 0.158$ , ref. 5;  $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ;  $q_0 = 0.5$ ). The jet is unresolved perpendicular to its axis, at least out to a distance from the core  $r = 30$  mas, and we assume that it is a continuous inward extension to the core D of the jet seen on a kiloparsec scale<sup>6</sup>. No counterjet is visible, and the formal ratio of jet to counterjet is about 300:1 beyond a few milliarcseconds from the core.

There are three main emission regions along the jet, denoted in Fig. 1a by R1, R2, and R3. Their peak intensities decrease with distance from the core and they are separated by distinct gaps where the intensity drops by a factor of 10 or more (Fig. 1b). The VLBI components<sup>1,2</sup> in 3C273 are not isolated and as well defined as, for example, those in 3C345 (ref. 7). Region R3 is  $\sim 4$  mas long, and must contain the components D, C8, and C7a seen at 10.7 GHz in 1985.6 (ref. 2), but these are not separately recognizable in Fig. 1a. R2 must contain components C5, C4, and C3. The motion of C5 has been followed since 1980 at 10.7 GHz (refs 1, 2) and the feature labelled C5 in Fig. 1a lies close to the position extrapolated from the 10.7 GHz epochs. We identify the elongated outer part of R2 with C3 and C4.

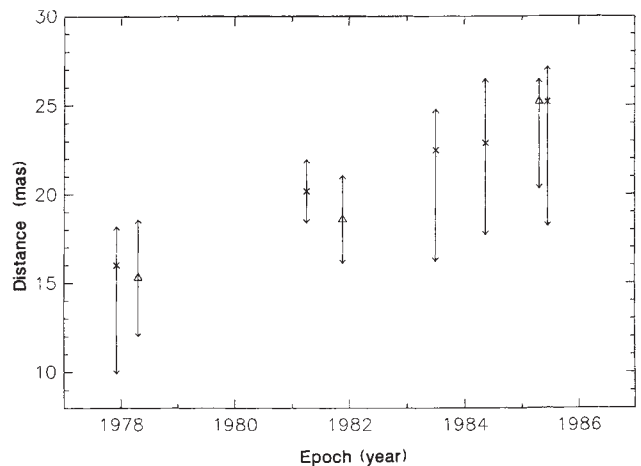


**Fig. 1** *a*, 5 GHz VLB1 map of 3C273 (epoch 1985.42), convolved with a restoring beam of  $2 \times 1$  mas in p.a.  $-11.7^\circ$ . Contour levels are  $-0.2, -0.1, 0.05, 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, 12.8, 25, 50$  and 90% of the peak, 8.37 Jy per beam. Typical r.m.s.-noise is  $\sim 2$  mJy per beam (except near sidelobe features) and the estimated thermal noise is 0.3 mJy per beam. This map was made from a full-track observation with the following stations: Toruń, Poland; Onsala, Sweden; Effelsberg, West Germany; Westerbork, Netherlands; Jodrell Bank, England; Medicina, Italy; Hartebeesthoek, South Africa; Arecibo, Puerto Rico; Haystack, Massachusetts; Green Bank, West Virginia; North Liberty, Iowa; VLA, New Mexico; Hat Creek and Big Pine, California. *b*, Inset: flux density (Jy) contained in bins of width 0.9 mas perpendicular to a line at p.a.  $-131^\circ$  (through the core and the ridge of C2) as a function of distance along jet milliarcseconds. The horizontal line is drawn at 1 mJy level and represents the one-sigma noise level of the map.

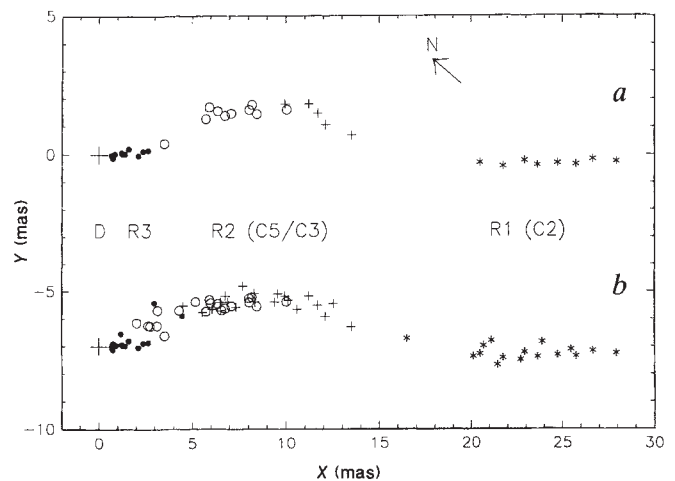
They were well separated in 1979 and 1980, but are partially blended in Fig. 1*a*, which is consistent with their somewhat different velocities<sup>1</sup>.

We identify region R1 with component C2 that has been seen at frequencies of 5 GHz and below since 1977<sup>1,8-10</sup>. It has always appeared extended, and the better maps show it as an elongated ridge with the brightest component at the leading edge. Figure 2 shows a graph of the separation from the core of the peak (and the inner and outer edges) of C2 as a function of time. This shows that C2 moves along the jet axis at a rate between 1.0 and 1.3 mas yr<sup>-1</sup>. The superluminal motions in 3C273 thus extend over the entire length of the well determined part of the jet, out to  $\sim 26$  mas, or  $l_{\text{proj}} \sim 46h^{-1}$  pc. The measured speeds have been  $\sim 1$  mas yr<sup>-1</sup>, without evidence for acceleration or deceleration.

A critical issue is the question of whether the motions are along the same or different paths, and whether these are 'ballistic' (namely straight lines that extend radially back to the core) or instead are curved and contain kinks. A decisive test is to follow individual components as they move to locations previously occupied by earlier components. In Fig. 3*a* we plot the locus of the ridgeline in the jet in 3C273. The points from  $r = 5-30$  mas are taken from Fig. 1, and to improve the accuracy for  $r < 5$  mas we added points from the 10.7 GHz maps at epochs between 1984.1 and 1985.6, which were also measured with high north-south resolution<sup>2</sup>. The plot shows that region R1 defines a radial line at position angle (p.a.)  $-131^\circ$ . R3 appears to start at about p.a.  $-135^\circ$  and bends towards p.a.  $-125^\circ$  at  $r = 3$  mas (ref. 2). R2 appears somewhat curved and is offset from the line



**Fig. 2** Locations of the leading peak of the C2 region (see text). Crosses are from data at 5 GHz: 1977.92 and 1981.26 (ref. 1), 1983.5 and 1984.3 from unpublished maps, and 1985.41 from this paper. Triangles are from data at 2.3 GHz: 1978.3 and 1981.89 (ref. 8), and 1985.37 (ref. 9). The vertical lines do not represent formal errors, but indicate the overall extent of C2. The length of the lines also depends on the quality of the maps. The early lines thus show lower limits to the overall extent, and there is some uncertainty about the location of the peaks.



**Fig. 3** Ridge line of the jet in 3C273, rotated counterclockwise by  $40^\circ$  (that is, a horizontal line has p.a.  $-130^\circ$  on the sky). *a*, Measured from Fig. 1*a* and ref. 2. *b*, As in *a*, with data from an unpublished 5 GHz map of epoch 1984.3 and additional points from ref. 1. Symbols: C2 (stars); C3 (crosses); C5 (open circles); all inner components (dots). Note that the data from refs 1 and 2 are only central positions of components measured of maps. The overlapping of the trajectories of C3 and C5 indicates that these two components have followed the same projected path.

R1-R3 by nearly 2 mas ( $3.5h^{-1}$  pc), defining a direction that does not intersect the origin. In Fig. 3*b* we added data from an unpublished 5 GHz map of epoch 1984.3, and the positions of components observed between 1977.9 and 1981.1 at 10.7 and 5 GHz (ref. 1). The latter were measured with poor north-south resolution and do not unambiguously show details of the ridge at  $r = 3-4$  mas. But Fig. 3*b* shows that the ridge is curved and suggests that it contains kinks at roughly  $r = 3-4$  mas and at  $r = 12-17$  mas from the core. We have followed C3 for over 6 years, covering  $r \approx 5-12$  mas. Its motion is roughly along p.a.  $-130^\circ$ , but offset from a radius; that is, C3 is not ballistic. The same holds for C5, which has moved from  $r \approx 2$  to  $r \approx 8$  mas



and is now following the track of C3. The tracks of C5 and C3 overlap for about 3 mas, and we assume that they have the same trajectory, at least in R2. There are no other well-defined overlaps, but the extended component C2 has been moving parallel to its length for at least 7 years. There is no evidence that this motion was not ballistic, although 3C273 has always shown curved structure<sup>11,12</sup>, suggesting that at earlier epochs C2 was also offset from a radial line, following the same path as C3 and C5. It will be crucial to determine whether C3 and C5 will follow the path of C2 during the next few years.

It may be significant that the kinks in the jet at  $r \approx 4$  and 15 occur at or near the gaps in surface brightness. The gaps themselves are not stationary, but move with the general flow. But, we do not know if the kinks are stationary, as we have not yet followed a component with sufficient resolution near a kink; this will be achieved by further observations at 10.7 GHz (ref. 2).

With a proper motion  $\mu \approx 1 \text{ mas yr}^{-1}$ , the apparent transverse velocity of the moving features in the reference frame of 3C273 is  $\beta_{\text{app}} = v_{\text{app}}/c \approx 7h^{-1}$  (ref. 13). On the usual relativistic interpretation, the motion is at an angle to the line of sight smaller than  $\theta_M \approx 16^\circ$ , and its Lorentz factor is greater than  $\gamma_M \approx 7h^{-1}$  (ref. 12). There is model-dependent evidence<sup>1</sup> that  $\theta \leq 5^\circ$ , because of the lack of an observed shift in the location of the 5 GHz and 10.7 GHz components. We assume that the jet has the smallest possible Lorentz factor  $\gamma = 7h^{-1}$ , which corresponds to  $\theta \approx 8^\circ$ ; with this assumption modest changes of  $\theta$  have only a small effect on  $\beta_{\text{app}}$  (ref. 14).

Our observations of 3C273 can be explained by a model in which a central machine produces a relativistic flow in a channel that has mild curvature maintained by suitable pressure gradients. The apparent change in direction by  $20^\circ$  is probably amplified and the underlying jet could be straight to within a few degrees. The absence of a counterjet is perhaps due to an intrinsic one-sidedness (3C273 is the only known object that appears one-sided on both small and large scales)<sup>6</sup>, but alternatively there could be an intrinsically strong counterjet, because beaming would reduce its intensity by a factor of order  $\gamma^4$  relative to the approaching jet. The dramatic variations in intensity along the jet (Fig. 1b) might be due to entrainment, or shock waves, or differential Doppler boosting; but it is also possible that they are caused by changes in output of the central machine. The reduction in strength of the successive peaks (Fig. 1b) may be due to lateral expansion. Then the projected opening angle of the jet would be less than  $4^\circ$  and the intrinsic angle would be  $1^\circ$  or less. This is consistent with the narrow intrinsic opening angle of  $0.2^\circ$  estimated for the extended jet<sup>15</sup>.

C2 appears as an enduring feature in the flow and its observed length is approximately  $w = 8 \text{ mas}$  or  $15 h^{-1} \text{ pc}$ . As it is moving relativistically its intrinsic length  $l_Q$  (in  $Q$ , the coordinate system of the quasar) is not simply given by a deprojection. Suppose the machine makes a constant-velocity feature for a time  $\tau_Q$ , then its length is  $l_Q = \beta c \tau_Q$  and  $w = \beta c \tau_Q \sin \theta (1 - \beta \cos \theta)^{-1} = \beta_{\text{app}} c \tau_Q$ . Thus  $l_Q = (\beta/\beta_{\text{app}}) w \sim 2 \text{ pc}$  and  $\tau_Q \sim 6 \text{ yr}$ . Both  $l_Q$  and  $\tau_Q$  are independent of the value of the Hubble constant for  $\gamma^2 \gg 1$ .

The structure of 3C273 is not readily explained by 'screen' models<sup>15</sup> in which the emitting material is stationary, and the moving patterns are reflections of a central variable source, or are the locations of the intersection of a moving jet with a screen. The simplest 'echo' models cannot explain the gap in Fig. 1a at  $r = 4 \text{ mas}$ , because there is no evidence for an interval of weak emission in the centre 4 years ago. More complicated narrow jet models<sup>16</sup> require several beams to produce the multiple moving components. Furthermore, the range in distance of  $>10:1$  over which superluminally-moving components have been seen is inconsistent with the modest reduction in amplitude along the jet (30:1), unless the particle density and magnetic field in the screen vary extremely slowly.

We thank the staff of the observatories participating in our experiment, P. Charlot and S. Unwin for making their maps

available before publication and P. Barthel, A. Readhead and D. Roberts for comments. This work was partly supported by the US NSF and by the Swedish NSRC.

Received 4 April; accepted 16 June 1988.

1. Unwin, S. C. *et al. Astrophys. J.* **289**, 109–119 (1985).
2. Cohen, M. H. *et al. Astrophys. J.* **315**, L89–L92 (1987).
3. Biretta, J. A. *et al. Astrophys. J.* **292**, L5–L8 (1985).
4. Bridle, A. H. & Perley, R. A. *Astr. Astrophys.* **22**, 319–358 (1984).
5. Schmidt, M. *Nature* **197**, 1040 (1963).
6. Davis, R. J., Muxlow, T. W. B. & Conway, R. G. *Nature* **318**, 343–345 (1985).
7. Biretta, J. A., Moore, R. L. & Cohen, M. H. *Astrophys. J.* **308**, 93–109 (1986).
8. Cohen, M. H. *et al. Astrophys. J.* **272**, 383–389 (1983).
9. Charlot, P., Lestrade, J.-F. & Boucher, F. *Proc. IAU Symp.* **129**, 33–34 (1988).
10. Unwin, S. C. & Davis, R. J. *Proc. IAU Symp.* **129**, 27–28 (1988).
11. Niell, A. E., Kellermann, K. I., Clark, B. G. & Shaffer, D. B. *Astrophys. J. Lett.* **197**, L109–L112 (1975).
12. Seielstad, G. A. *et al. Astrophys. J.* **229**, 53–72 (1979).
13. Zensus, J. A. & Pearson, T. J. (eds) *Superluminal Radio Sources* 1–11 (Cambridge University Press, 1987).
14. Unwin, S. C. *et al. Astrophys. J.* **271**, 536–550 (1983).
15. Conway, R. G., Davis, R. J. & Foley, A. R. *Nature* **294**, 540–542 (1981).
16. de Waard, G. J. thesis, Leiden Univ. (1986).

## Supernova 1987A: radiosphere resolved with VLBI five days after the neutrino burst

D. L. Jauncey\*, A. Kembell†, N. Bartel‡, A. R. Whitney§, A. E. E. Rogers§, I. I. Shapiro‡, R. A. Preston||, T. A. Clark¶, B. R. Harvey\*, D. L. Jones\*, G. D. Nicolson†, A. Nothnagel†, R. B. Phillips§, J. E. Reynolds\*\* & J. C. Webber§

\* CSIRO, Division of Radiophysics, PO Box 76, Epping, New South Wales 2121, Australia

† Hartebeesthoek Radio Astronomy Observatory, PO Box 3718, Johannesburg 2000, South Africa

‡ Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

§ Haystack Observatory, NEROC, Westford, Massachusetts 01886, USA

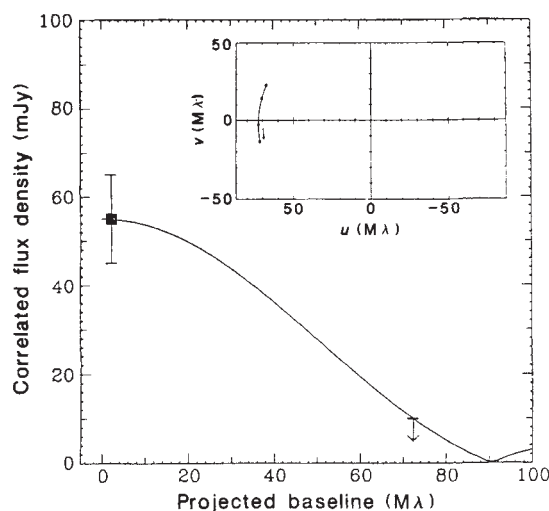
|| Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA

¶ NASA, Goddard Space Flight Center, Code 974, Greenbelt, MD 20771, USA

\*\* Mount Stromlo and Siding Springs Observatory, Private Bag, PO Woden, ACT 2606, Australia

Following the detection<sup>1</sup> of radio emission from SN1987A in the Large Magellanic Cloud, we conducted very-long-baseline interferometry (VLBI) observations with an interferometer composed of a NASA Deep Space Network antenna (DSS42) at Tidbinbilla, Australia and the antenna of the Hartebeesthoek Radio Astronomy Observatory, South Africa<sup>2,3</sup>. We did not detect any emission from the supernova above a level of  $\sim 20\%$  of the supernova's total flux density, although signals were detected from our two calibrator sources with amplitudes roughly equal to those determined in earlier VLBI observations. We infer that we resolved the supernova's radiosphere and estimate, for an epoch 5.2 days after the neutrino burst<sup>4,5</sup>, a lower bound on the radiosphere's radius of 1.2 mas. Given the photometric data from the supernova<sup>6,7</sup>, a distance to the Large Magellanic Cloud of  $50 \pm 5 \text{ kpc}$  (ref. 8), and an apparent expansion velocity that varied systematically with time from  $18\text{--}16 \times 10^3 \text{ km s}^{-1}$  (refs 9 and 10), as estimated from the blue-shifted H $\alpha$  absorption lines on the days preceding our observations, we conclude that 5.2 days after the neutrino burst the supernova's radiosphere was at least 2.5 times larger than the inferred blackbody photosphere, and at least as large as the H $\alpha$  line-forming region.

The discovery of a supernova in a galaxy like the Large Magellanic Cloud, to which the distance of  $50 \pm 5 \text{ kpc}$  is known



**Fig. 1** The total flux density<sup>1</sup> (square) and an upper bound on the correlated flux density of SN1987A on ~28.5 February 1987. The upper bound was obtained from incoherently adding the data from four 10-min scans on the supernova (see text). The corresponding  $u$ - $v$  track for these observations, from 1130 UT to 1400 UT, is plotted in the inset; the projected baseline varied by less than 10% in length. The dots correspond to the centres of the four scans. The solid curve represents the prediction of a model of an optically thin uniform sphere with a radius of 1.6 mas, matched to the total flux density and our upper bound on the correlated flux density. This prediction is similar to that from a model of an optically thin shell with an outer diameter of 1.25 mas and a shell thickness of about 15%.

rather precisely, and the subsequent report of the detection of radio emission<sup>1</sup>, provided an unprecedented opportunity: we could determine the linear radius, or an upper bound on it, of the supernova's radiosphere from VLBI observations. Such an estimate could then be compared with the linear radii of the supernova's photosphere and line-forming regions. A comparison is important in view of the partial conflict between predictions from theoretical models<sup>11-13</sup> and the possibility of estimating distances to supernovae by dividing the linear radius of the supernova, inferred from optical spectra<sup>14-16</sup>, by the angular radius of a supernova's radiosphere determined with VLBI.

Using the largest observed expansion velocity of  $31 \times 10^3 \text{ km s}^{-1}$  from the blue edge of the  $H\alpha$  absorption trough on 25 February 1987 (ref. 9), we inferred that the rate of increase of the radius of the radiosphere could be as large as  $0.4 \text{ mas d}^{-1}$ . The only interferometer with the recording equipment, sensitivity, and baseline length adequate for at least partly resolving the supernova a few days after the neutrino burst<sup>4,5,17</sup>, was any one of the antennae of the NASA Tidbinbilla complex in Australia and the antenna in Hartebeesthoek in South Africa. The latter operated as part of the IRIS-S geodetic VLBI network at the time of the supernova explosion. We therefore used this capability to make VLBI observations of SN1987A in right circular polarization at 2.3 GHz on each of the days 28 February and 1, 2 March 1987. We used as one element of the interferometer the DSS42 antenna at Tidbinbilla, which has a diameter of 34 m and a sensitivity of  $0.192 \text{ kJy}^{-1}$ , and provides a system temperature of ~21 K with its maser receiver. The antenna in Hartebeesthoek has a diameter of 26 m and provides, with a temporary feed and uncooled FET amplifier, a sensitivity of  $0.053 \text{ kJy}^{-1}$  and a system temperature of ~85 K.

The observations of SN1987A were alternated with those of known compact sources, mostly of 0537-441 and 0521-365, and lasted for 2.5-3.5 hours on each day. The  $u$ - $v$  track of the first day is shown as an inset in Fig. 1. The fringe spacing was nearly constant at about 2.9 mas. A hydrogen maser was used as a time

and frequency standard at each station. All observations were made with the Mark III VLBI system in recording mode A (ref. 18). A bandwidth of 48 MHz, centred at about 2,278 MHz, was recorded at each site, with an offset of 2 MHz having been inadvertently introduced between the bands recorded at one site and that at the other. The data were correlated with the Mark III A processor at the Haystack Observatory. A specially implemented (time-consuming) mode had to be used to correct for the effects of the 2 MHz offset. This mode was checked with observations of 3C84 made on 21 March 1987 with an interferometer composed of the antennae at Richmond, Florida, and Westford, Massachusetts. As expected, no amplitude degradation was found between the data recorded with an offset of 2 MHz and processed in the special mode, and the data recorded with no offset and processed in the normal mode.

As the coordinates of DSS42 were not known with sufficient accuracy for a confined, and hence more sensitive, search for fringes from the supernova's radio emission, we conducted additional VLBI observations. These were made on 8 and 9 June 1987 with an interferometer composed of, again, the antenna at Hartebeesthoek and, this time, the DSS45 antenna at Tidbinbilla. From our solution for the coordinates of DSS45 and knowledge of the baseline vector from DSS45 to DSS42 from a previous ground survey of the Tidbinbilla station complex, we determined the coordinates of DSS42 with sufficient accuracy for a sensitive fringe search<sup>19</sup>.

We confined our search for fringes from the supernova to an area on the sky of  $\pm 0.8$  arcsec in right ascension and declination, centred at the optical position<sup>20</sup>  $\alpha(\text{B1950.0}) = 05 \text{ h } 35 \text{ min } 49.95 \text{ s}$  and  $\delta(\text{B1950.0}) = 69^\circ 17' 57.9''$  of the supernova in the Perth system. This search area is large enough to encompass the uncertainty of the optical position determination of  $< 0.2$  arcsec in both coordinates<sup>20-22</sup> and any offset of the optical reference frame from the radio reference frame of  $< 0.6$  arcsec (refs 19 and 23).

No fringes were found from the supernova's radio emission on any of the three observing days. We calibrated the peaks observed in the correlation coefficients in the usual manner for Mark III VLBI data (see ref. 24, for example), incoherently averaged the data for each day, and obtained (putative) correlated flux densities up to 9.7 mJy on 28 February, up to 10.5 mJy on 1 March and up to 11.3 mJy on 2 March. The peak on the first day was located 0.7 arcsec away from the expected position of the supernova and the peaks on the following days were widely scattered over the search area. The largest peak in an area 0.3 arcsec around the expected position corresponded on 28 February to a correlated flux density of only 7 mJy. Given the standard errors of our calibration procedures, we conclude that there were no signals from the supernova on 28 February with a correlated flux density exceeding  $10 \pm 1$  mJy and on each of the following days with a correlated flux density exceeding  $11 \pm 1$  mJy. Future improvements in the radio-optical registration or a future direct determination of the radio position of the supernova remnant, should the remnant become detectable again, may allow reducing the search area and thus improving the upper limits on the correlated flux density (see ref. 3 for a review of radio observations of SN1987A).

Fringes were obtained from all observations of both of the compact calibrator sources; the corresponding correlated flux densities were 2.3 Jy for 0537-441 and 0.44 Jy for 0521-365, with estimated standard errors of ~10% for each source. These results agree, to within the root-sum-squares of their estimated standard errors, with VLBI measurements made five years earlier at the same resolution (ref. 25 and G. D. Nicolson, personal communication). This agreement may be coincidental at least in part, as the second calibrator source is known to vary by more than the difference in these results<sup>25</sup>.

In Fig. 1 we compare our upper bound on the supernova's correlated flux density of  $10 \pm 1$  mJy (uncertainty not plotted) on 28 February with the corresponding (interpolated) total flux



density of  $55 \pm 10$  mJy at the midtime of our VLBI observations. The latter value was obtained from an interpolation of total flux density values<sup>1</sup> determined with the Parkes Tidbinbilla interferometer (PTI, baseline length: 275 km)<sup>26</sup> at three epochs (one day before,  $\sim 3$  h before, and one day after our VLBI observations), and from a subsequent multiplication by a factor of 0.85 resulting from a recalibration of the PTI data (R. P. Norris, personal communication, and ref. 3).

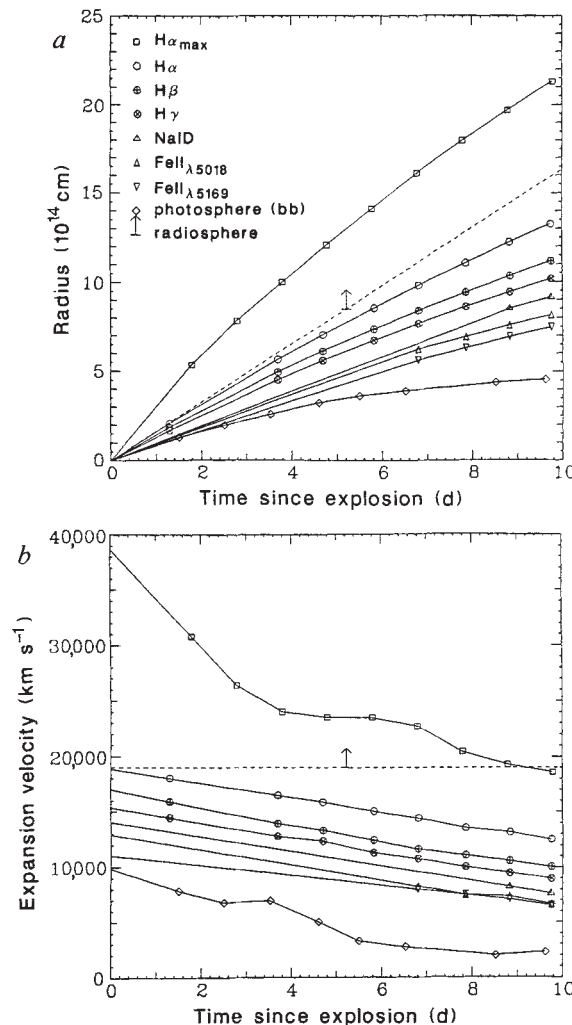
To determine a lower bound on the angular diameter of the supernova's radiosphere, we considered a range of models for the radiosphere's brightness distribution. In Fig. 1 we match the predictions from an optically thin uniform sphere model to the total flux density and to the upper bound on the correlated flux density; this match yields a lower-bound estimate on the sphere's radius of  $1.6 \pm 0.1$  mas. An optically thick sphere or shell model yields a lower bound on the radius  $\sim 10\%$  smaller than that for the optically thin uniform sphere model<sup>27</sup>. But as the supernova's radio spectrum had a spectral index,  $\alpha$  ( $S_\nu \propto \nu^\alpha$ ), between 0.84 and 2.3 GHz of about  $-0.4$  to  $-0.6$  on 28 February 1987 (ref. 1), indicating that the source was optically thin, we ignore optically thick models. A model of a thin circular ring oriented normal to our line of sight will lead to the smallest inferred angular radius<sup>27</sup> of only  $1.00 \pm 0.05$  mas. But we do not consider such a model physically plausible and instead we use an optically thin shell model with a shell thickness of  $\sim 15\%$  of the shell's outer diameter to derive a lower bound on the radiosphere's (outer) angular radius of  $1.25 \pm 0.07$  mas. The use of a model with a smaller shell thickness has only a marginal effect on our bound, smaller than the bound's standard error. Given a distance to SN1987A of  $50 \pm 5$  kpc (ref. 8), our lower bound on the angular radius corresponds to a lower bound on the linear radius of  $R_{\text{radio}} > 8.3 \times 10^{14}$  cm  $\approx 12 \times 10^3 R_\odot \approx 55$  AU, 5.2 days after the neutrino burst<sup>4,5</sup>. If we assume that the radiosphere expanded linearly from a zero size at the epoch of the neutrino burst, our lower bound on the radius corresponds to a lower bound on the expansion velocity of the radiosphere of  $v_{\text{radio}} > 19 \times 10^3$  km s<sup>-1</sup> (see Fig. 2).

How do our lower bounds on the radius and expansion velocity of the supernova's radiosphere compare with the radii and expansion velocities of the supernova's photosphere and line-forming regions? Menzies *et al.*<sup>6</sup> and Catchpole *et al.*<sup>7</sup> were able to fit absorption-corrected photometric data extremely well with a blackbody spectrum and to derive the photosphere's angular radius,  $\Theta_{\text{bb}}$ , as a function of time. Other investigators determined the blueshifts of the FeII $_{\lambda 5169}$ , FeII $_{\lambda 5018}$ , NaID, H $\delta$ , H $\beta$  and H $\alpha$  absorption minima<sup>10</sup> and the largest observed blueshift from the blue edges of the H $\alpha$  absorption trough<sup>9</sup>, to determine the equivalent respective expansion velocities of the corresponding supernova gas.

In Fig. 2a we compare our lower bound on the radiosphere's linear radius with the linear radii of the photosphere and of the metal and hydrogen gas clouds, the latter obtained from an integration of the corresponding velocities. In Fig. 2b we compare our lower bound on the radiosphere's expansion velocity with the expansion velocity of the photosphere, obtained from a differentiation of  $R_{\text{bb}}$ , and with the velocity values for the other gases.

For any given epoch, the photosphere has the smallest radius and expansion velocity. The photosphere is surrounded by a line-radiation absorbing atmosphere containing a metal and hydrogen gas with an apparently radially decreasing opacity. At the time of our VLBI observations, the radius and expansion velocity of the radiosphere were both at least a factor of 2.5 larger than those of the photosphere and at least 10% and 25% larger than  $R_{\text{H}\alpha}$  and  $v_{\text{H}\alpha}$  respectively.

Our results are consistent with models of supernovae in which radio emission originates in a magnetosphere outside the photosphere<sup>11</sup>, or more specifically, in the interaction region between the circumstellar medium and the supernova shockfront<sup>12</sup>. Such a supernova shockfront could expand with a velocity equal to



**Fig. 2** *a*, Lower bound on the radius of the radiosphere compared with values for the radii of the photosphere and of the line-forming region. The latter radii were obtained from an integration of the velocities corresponding to the blueshifts of the absorption of the indicated lines and the largest observed blueshift, that from the blue edge of the H $\alpha$  absorption trough. The dashed line represents a uniform expansion with a speed of  $19 \times 10^3$  km s<sup>-1</sup>, our lower bound on the expansion velocity of the radiosphere. In *a* and *b*, the supernova's distance is assumed to be 50 kpc. The 5 kpc uncertainty in this distance is included in the determinations of the lower bounds on the radiosphere's radius, and hence on its velocity, but is not indicated in the equivalent values of the photosphere. *b*, Lower bound on the average expansion velocity of the radiosphere is compared with the expansion velocities of the photosphere, obtained from differentiation of the values in *a*, and with the expansion velocities of the line-forming regions. The meaning of the symbols is given in *a*. The uncertainties of the H $\alpha_{\text{max}}$  velocities are about  $\pm 1,000$  km s<sup>-1</sup> and those of the H $\alpha_{\text{max}}$  radii are smaller than the symbols. Other uncertainties were not given in the original papers but are believed to be not larger than those of the corresponding H $\alpha_{\text{max}}$  velocities and radii.

$v_{\text{max}}$ . The outer diameter of the radiosphere 5.2 days after the neutrino burst could therefore be  $\sim 50\%$  larger than our lower bound on it (see also ref. 28). As the distance to the Large Magellanic Cloud is known to within  $\sim 10\%$ , SN1987A allowed a comparison to be made for the first time of the linear radii and expansion velocities of a supernova's photosphere and radiosphere with the supernova's line-forming regions (see also ref. 29 for results from speckle observations). Apart from an improvement of our understanding of supernovae, such a com-

parison should also allow a more reliable use of supernovae as distance indicators<sup>30-32</sup>, if the physical processes responsible for the emission from SN1987A are typical for supernovae in general<sup>3,7,33,34</sup>.

We thank Tom Reid, Paul Hutchinson, and the staff of the Deep Space Communications Complex, Tidbinbilla, for their generous assistance and Turtle *et al.* for making their results available to us before publication. Research at the Harvard-Smithsonian Center for Astrophysics and at the Haystack Observatory was supported in part by the US National Science Foundation (NSF). Research at the Jet Propulsion Laboratory, California Institute of Technology, was carried out under contract to the US National Aeronautics and Space Administration (NASA). The support of NASA's Deep Space Network staff at Tidbinbilla, Australia, and the help of the US National Oceanic and Atmospheric Administration for arranging the temporal operation of a receiver and a Mark III data acquisition system for observations at Hartebeesthoek, South Africa, are greatly appreciated.

Received 6 June; accepted 20 June 1988.

1. Turtle, A. J. *et al.* *Nature* **327**, 38-40 (1987).
2. Shapiro, I. I. *et al.* in *IAU Symp. No. 129, The Impact of VLBI on Astrophysics and Geophysics* (eds Reid, M. J. & Moran, J. M.) 185-186 (Reidel, Dordrecht, 1988).
3. Bartel, N. *et al.* in *Supernova 1987A in the Large Magellanic Cloud* (eds Kafatos, M. & Michalitsianos, A.) 81-86 (Cambridge University Press, 1988).
4. Hirata, K. *et al.* *Phys. Rev. Lett.* **58**, 1490-1493 (1987).
5. Bionta, R. M. *et al.* *Phys. Rev. Lett.* **58**, 1494-1496 (1987).
6. Menzies, J. W. *et al.* *Mon. Not. R. astr. Soc.* **227**, 39P-49P (1987).

7. Catchpole, R. M. *et al.* *Mon. Not. R. astr. Soc.* **229**, 15P-25P (1987).
8. Feast, M. W. & Walker, A. R. *Astr. Astrophys.* **25**, 345-375 (1987).
9. Hanuschik, R. W. & Dachs, J. *Astr. Astrophys. Lett.* **182**, L29-L30 (1987).
10. Blanco, V. M. *et al.* *Astrophys. J.* **320**, 589-596 (1987).
11. Shklovskii, I. S. *Pis'ma Astr. Zh.* **7**, 479-481 (1981), translation *Sov. Astr. Lett.* **7**, 263-264 (1981).
12. Chevalier, R. A. *Astrophys. J.* **259**, 302-310 (1982).
13. Pacini, F. & Salvati, M. *Astrophys. J.* **245**, L107-L108 (1981).
14. Bartel, N. in *Supernovae as Distance Indicators, Lecture Notes in Physics* Vol. 224 (ed. Bartel, N.) 107-122 (Springer, Berlin, 1985).
15. Bartel, N. *et al.* *Nature* **318**, 25-30 (1985).
16. Bartel, N. in *IAU Symp. No. 129, The Impact of VLBI on Astrophysics and Geophysics* (eds Reid, M. J. & Moran, J. M.) 175-184 (Reidel, Dordrecht, 1988).
17. Aglietta, M. *et al.* *Europhys. Lett.* **3**, 1315-1320 (1987).
18. Rogers, A. E. E. *et al.* *Science* **219**, 51-54 (1983).
19. Harvey, B. R. *et al.* *Astronomical J.*, to be submitted.
20. White, G. L. & Malin, D. F. *Nature* **327**, 36-38 (1987).
21. West, R. M., Lamberts, A., Jørgensen, H. E. & Schuster, H.-E. *Astr. Astrophys.* **177**, L1-L3 (1987).
22. Girard, T., van Altena, W. F. & Lopez, C. E. *Astronomical J.* **95**, 58-63 (1988).
23. White, G. L., Jauncey, D. L. & Preston, R. A. *Astronomical J.*, submitted.
24. Bartel, N. *et al.* *Astrophys. J.* **262**, 556-563 (1982).
25. Preston, R. A. *et al.* *Astrophys. J.* **90**, 1599-1641 (1985).
26. Norris, R. P., Batty, M. J., Kesteven, M. J. & Wellington, K. J. in *IAU Symp. No. 129, The Impact of VLBI on Astrophysics and Geophysics* (eds Reid, M. J. & Moran, J. M.) 487-488 (Reidel, Dordrecht, 1988).
27. Marscher, A. P. in *Supernovae as Distance Indicators, Lecture Notes in Physics* Vol. 224 (ed. Bartel, N.) 130-137 (Springer, Berlin, 1985).
28. Chevalier, R. A. & Fransson, C. *Nature* **328**, 44-45 (1987).
29. Karovska, M., Nisenson, P., Papaliolios, C. & Standley, C. *IAU Circ.* **4457** (1987).
30. Branch, D. *Astrophys. J. Lett.* **320**, L23-L25 (1987).
31. Bartel, N. (ed.) *Supernovae as Distance Indicators, Lecture Notes in Physics* Vol. 227 (Springer, Berlin, 1985).
32. Bartel, N. *Highlights of Astronomy* Vol. 7 (ed. Swings, J. P.) 655-660 (Reidel, Dordrecht, 1986).
33. Storey, M. C. & Manchester, R. N. *Nature* **329**, 421-432 (1987).
34. Weiler, K. W., Sramek, R. A., Panagia, N., van der Hulst, J. M. & Salvati, M. *Astrophys. J.* **301**, 790-812 (1986).

## Contribution of Cl- and F-bearing gases to the atmosphere by volcanoes

Robert B. Symonds, William I. Rose & Mark H. Reed\*

Department of Geological Engineering, Geology & Geophysics, Michigan Technological University, Houghton, Michigan 49931, USA

\* Department of Geological Sciences, University of Oregon, Eugene, Oregon 97403, USA

As halogen gases catalyse the destruction of ozone in the stratosphere<sup>1-3</sup>, it is important to quantify the natural emissions of halogens from active volcanoes. Here we use equilibrium thermodynamics to predict the speciation of Cl and F in volcanic gases and provide new estimates of the global emission rates to the atmosphere. Our calculations show that HCl and HF are the dominant species of Cl and F in volcanic gases, at least several orders of magnitude more abundant than all other species. We estimate the annual global volcanic fluxes of HCl and HF to be 0.4-11 Tg ( $10^{12}$  g) and 0.06-6 Tg, respectively. On average, <10% of these emissions come in large explosive eruptions that transmit them efficiently to the stratosphere. Although they are infrequent, large volcanic eruptions may inject significant amounts of HCl and HF into the stratosphere. Passively degassing volcanoes are also a major source of tropospheric HF, although sea salt is the main source of tropospheric HCl.

Continued production of industrial halocarbons may lead to the destruction of the ozone layer<sup>1</sup> and contribute to greenhouse warming<sup>2</sup>. Many halogen gases have been shown to destroy ozone in the laboratory, but most are removed by tropospheric processes (precipitation, reactions) before they reach the ozone layer in the stratosphere<sup>3</sup>. Halocarbons are unusual among halogens because their atmospheric residence times are generally long enough to cause their dispersion to the stratosphere, where they participate in photochemical reactions that destroy ozone<sup>3</sup>. With the exception of methyl halides (mostly  $\text{CH}_3\text{Cl}$ ), which are released from the oceans and by burning vegetation, industrial emissions are apparently the dominant source of most

halocarbons<sup>3</sup>. Between 1982 and 1984, the annual global industrial production<sup>3</sup> of Cl and F as halocarbons was 2.28 Tg and 0.273 Tg, respectively. The annual release of Cl as  $\text{CH}_3\text{Cl}$  from the oceans and burning amounts to 1.4-3.5 Tg (ref. 3), most of which comes from the oceans<sup>4-5</sup>.

Active volcanoes continuously release significant quantities of halogens, as they have throughout Earth history. The dominance of chloride among anions in the oceans has been shown to reflect the volcanic degassing of Cl-bearing gases over aeons<sup>6-7</sup> and the solubility of many Cl-bearing gases in  $\text{H}_2\text{O}$ . Given that volcanoes release halogens, it is important to estimate the natural fluxes and to determine the molecular form of halogens in these emissions.

Volcanoes are one source of stratospheric chlorine<sup>8</sup>. The 1976 eruption of Augustine volcano injected 0.08-0.18 Tg of HCl into the stratosphere<sup>9</sup>. An estimated 40% increase in the stratospheric HCl between 20° and 40° N latitude was observed in 1982 (ref. 10), together with stratospheric halite ( $\text{NaCl}$ ) particles<sup>11</sup>. This was attributed to an estimated 0.04 Tg of HCl contributed to the stratosphere from the El Chichón eruption<sup>10</sup>.

The main volcanic gas species ( $\text{H}_2\text{O}$ ,  $\text{H}_2$ ,  $\text{CO}_2$ ,  $\text{CO}$ ,  $\text{SO}_2$ ,  $\text{H}_2\text{S}$ ,  $\text{S}_2$ , HCl, HF) can be analysed in small samples collected from volcanic fumaroles<sup>12</sup>. Typically, all the Cl and F in the analyses are assigned to HCl and HF (which cannot be analysed directly)<sup>12</sup>, even though trace amounts of metal halides<sup>13</sup> and chlorocarbons<sup>14-16</sup> occur in volcanic gases. Thus, it is important to know the relative proportions of various halogen gas species emitted from volcanoes.

To estimate the concentrations of many possible trace gas species, we used program SOLVGAS, a computer code for calculating homogeneous equilibrium in ideal gases<sup>13</sup>. The program calculates the distribution of hundreds of gas species in a system of 30-40 components as a function of temperature and pressure using the basic formulations of equilibrium calculations for aqueous systems<sup>17</sup> and modified for gases. Thermochemical data<sup>13</sup> constrain the calculations, which depend on the simultaneous solution of a series of mass balance and mass action equations using a Newton-Raphson method.

We applied SOLVGAS to volcanic gases from Mount St Helens, Kilauea, White Island, Mount Etna and Merapi vol-



**Table 1** Calculated fugacities of Cl and F gas species (for the C-O-H-S-Cl-F system) in the June 1980 volcanic gas of Mount St Helens

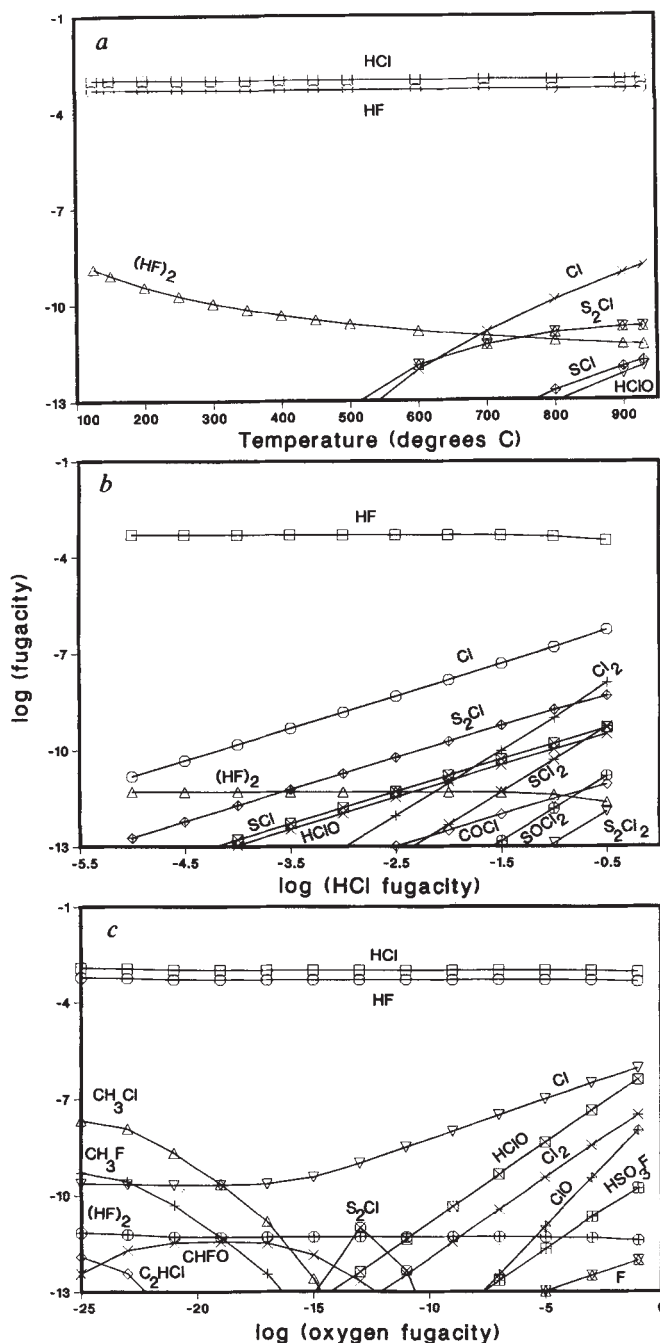
| Gas species                     | log fugacity | Gas species                     | log fugacity |
|---------------------------------|--------------|---------------------------------|--------------|
| HCl                             | -3.0         | ClF                             | -18.6        |
| HF                              | -3.3         | ClF <sub>3</sub>                | -50.1        |
| CCl                             | -28.7        | ClF <sub>5</sub>                | <-70.0       |
| CClF <sub>3</sub>               | -32.6        | ClF <sub>5</sub> S              | -54.5        |
| CCl <sub>2</sub>                | -27.5        | ClO                             | -16.4        |
| CCl <sub>2</sub> F <sub>2</sub> | -33.5        | ClO <sub>2</sub>                | -26.4        |
| CCl <sub>3</sub>                | -31.2        | ClO <sub>3</sub> F              | -51.0        |
| CCl <sub>3</sub> F              | -34.3        | Cl <sub>2</sub>                 | -13.0        |
| CCl <sub>4</sub>                | -34.8        | Cl <sub>2</sub> O               | -25.6        |
| CF                              | -26.1        | F                               | -14.8        |
| CF <sub>2</sub>                 | -25.6        | FO                              | -24.6        |
| CF <sub>3</sub>                 | -31.7        | FO <sub>2</sub>                 | -29.7        |
| CF <sub>3</sub> OF              | -47.1        | FS <sub>2</sub> F               | -19.8        |
| CF <sub>4</sub>                 | -31.6        | F <sub>2</sub>                  | -29.1        |
| CHCl                            | -25.4        | F <sub>2</sub> O                | -39.2        |
| CHClF <sub>2</sub>              | -27.3        | HCIO                            | -12.0        |
| CHCl <sub>2</sub> F             | -27.8        | (HF) <sub>2</sub>               | -11.3        |
| CHCl <sub>3</sub>               | -27.9        | (HF) <sub>3</sub>               | -19.3        |
| CHF                             | -24.5        | (HF) <sub>4</sub>               | -27.0        |
| CHFO                            | -13.1        | (HF) <sub>5</sub>               | -34.6        |
| CHF <sub>3</sub>                | -26.7        | (HF) <sub>6</sub>               | -41.8        |
| CH <sub>2</sub> ClF             | -22.5        | (HF) <sub>7</sub>               | -50.0        |
| CH <sub>2</sub> Cl <sub>2</sub> | -21.8        | HSO <sub>3</sub> F              | -15.5        |
| CH <sub>2</sub> F <sub>2</sub>  | -22.8        | SCl                             | -11.8        |
| CH <sub>3</sub> Cl              | -16.5        | SCl <sub>2</sub>                | -14.3        |
| CH <sub>3</sub> F               | -18.1        | SF                              | -13.7        |
| COCl                            | -13.5        | SF <sub>2</sub>                 | -18.5        |
| COCIF                           | -17.0        | SF <sub>3</sub>                 | -28.3        |
| COCl <sub>2</sub>               | -18.1        | SF <sub>4</sub>                 | -35.5        |
| COF                             | -17.3        | SF <sub>5</sub>                 | -48.7        |
| COF <sub>2</sub>                | -16.4        | SF <sub>6</sub>                 | -55.7        |
| CSF <sub>8</sub>                | <-70.0       | SOCl <sub>2</sub>               | -15.8        |
| C <sub>2</sub> Cl <sub>2</sub>  | -32.0        | SOF <sub>2</sub>                | -18.0        |
| C <sub>2</sub> Cl <sub>4</sub>  | -43.2        | SO <sub>2</sub> ClF             | -20.2        |
| C <sub>2</sub> Cl <sub>6</sub>  | -59.4        | SO <sub>2</sub> Cl <sub>2</sub> | -20.9        |
| C <sub>2</sub> F <sub>2</sub>   | -40.3        | SO <sub>2</sub> F <sub>2</sub>  | -20.0        |
| C <sub>2</sub> F <sub>4</sub>   | -47.7        | SSF <sub>2</sub>                | -17.1        |
| C <sub>2</sub> F <sub>6</sub>   | -55.8        | S <sub>2</sub> Cl               | -10.7        |
| C <sub>2</sub> HCl              | -25.9        | S <sub>2</sub> Cl <sub>2</sub>  | -16.9        |
| C <sub>2</sub> HF               | -30.1        | S <sub>2</sub> F <sub>10</sub>  | <-70.0       |
| Cl                              | -8.8         |                                 |              |

log fugacities in bars, assuming 1 bar total pressure.

canoes. We present the results from Mount St Helens, which are similar to those for the other volcanoes. We assumed the gas composition extrapolated to mid-June 1980 (ref. 18), a magmatic temperature of about 930 °C (ref. 19), and a total pressure of 1 bar in order to approach the gas erupted during the 18 May 1980 eruption. At first we only consider the gas species in the C-O-H-S-Cl-F system.

The calculated fugacities (partial pressures) demonstrate that HCl and HF are the dominant gas species of Cl and F by factors of at least a million (Table 1). Halocarbon fluxes are at least 13 orders of magnitude smaller than HCl and HF, which shows that they are probably extremely small. For instance, the 18 May 1980 eruption of Mount St Helens would have produced only 0.03 g of CH<sub>3</sub>Cl (using the concentration in Table 1 and SO<sub>2</sub> flux in Table 2), the most abundant halocarbon species. Assuming this same gas composition applies to all volcanoes, and an average global SO<sub>2</sub> flux of 18.7 Tg yr<sup>-1</sup> (ref. 20), the annual flux of CH<sub>3</sub>Cl is <1 g. Even using the 350-times-larger SO<sub>2</sub> emission for the Toba eruption 75,000 years ago (Table 2), the production of CH<sub>3</sub>Cl would be many orders of magnitude smaller than the HCl flux (Table 2).

To test whether there are conditions for which other halogen gas species become dominant, we repeated the calculations, varying temperature, HCl fugacity and oxygen fugacity. On cooling of the Mount St Helens gas, HCl and HF remain



**Fig. 1** Most stable gas species of Cl and F (in the C-O-H-S-Cl-F system) for the Mount St Helens volcanic gas (see text) as a function of *a*, temperature, *b*, HCl fugacity and *c*, oxygen fugacity. Plots *b* and *c* are at the assumed magmatic temperature of 930 °C. The HCl and oxygen fugacities in plots *b* and *c* extend well above and below conditions of typical volcanic gases. For instance, at 930 °C, the calculated logs of the HCl and oxygen fugacities for the Mount St Helens gas are about -3.0 and -12.2, respectively.

dominant, while most of the other halogen species decrease in abundance (Fig. 1).

Increasing HCl fugacity results in larger concentrations of Cl-bearing species as one would expect (Fig. 1), but even under geologically extreme conditions (30 mol% HCl), halocarbons and other Cl-bearing species still have very small concentrations compared to HCl.

At oxygen fugacities that are much smaller or larger than in typical volcanic fumaroles, HCl and HF remain the dominant species of Cl and F (Fig. 1). The equilibrium concentrations of

**Table 2** Volcanic contributions of HCl and HF to the atmosphere

| Explosive eruptions           |                     |                    |      |
|-------------------------------|---------------------|--------------------|------|
| Volcano and date              | HCl (Tg)            | HF (Tg)            | Ref. |
| Toba, 75,000 yr BP            | 400–4,000           | 20–2,000           | *    |
| Tambora, 1815                 | 216                 | 126                | 26   |
| Krakatau, 1883                | 3.75                | ND                 | 26   |
| St. Helens, 5/80              | 0.67                | 0.18               | †    |
| St. Helens 5/80               | 0.0352              | 0.053              | 26   |
| Augustine 1976                | 0.08–0.18           | ND                 | 9    |
| Augustine, 1986‡              | 0.07–0.9            | ND                 | 24   |
| El Chichón, 1982              | 0.04                | ND                 | 10   |
| Soufriere, 1979               | 0.035               | ND                 | 26   |
| Nonexplosive eruptions        |                     |                    |      |
| Volcano and date              |                     |                    |      |
| Laki, 1783                    | 1.6                 | ND                 | 26   |
| Agung, 1963                   | 1.53                | 0.8                | 26   |
| Hekla, 2,800 yr BP            | 0.26                | ND                 | 25   |
| Surtsey, 1963                 | 0.14                | ND                 | 25   |
| Krafla, 1981                  | 0.0038              | ND                 | 25   |
| Kilauea, 1984                 | 0.0003              | 0.0002             | 33   |
| Passively degassing volcanoes |                     |                    |      |
| Volcano and date              | HCl (Tg)<br>per day | HF (Tg)<br>per day |      |
| Augustine, 7/86               | 0.008               | ND                 | 24   |
| Etna, 1976                    | 0.00031             | ND                 | 34   |
| White Island, 1983            | 0.00042             | 0.00013            | 23   |
| Poás, 2/82                    | 0.00004             | 0.000003           | 35   |
| Merapi, January, 1984         | 0.00003             | 0.000002           | 13   |
| St Helens, 9/80               | 0.000004            | ND                 | 36   |

ND, no data or not detected.

\* Assuming 6,500 Tg of SO<sub>2</sub> (C. A. Chesner, personal communication); SO<sub>2</sub>/HCl, 1–10 and SO<sub>2</sub>/HF, 1–100 (see text).

† Assuming an SO<sub>2</sub> emission of 6 Tg (ref. 32); SO<sub>2</sub>/HCl, 5 and SO<sub>2</sub>/HF, 10 (calculated ratios at 930 °C for gas composition described in text).

‡ Minimum estimate from direct measurements and maximum calculated assuming an eruption equivalent to the 1976 Augustine eruption (0.186 km<sup>3</sup> dense-rock equivalent volume).

halocarbons generally increase with decreasing oxygen fugacity, but are very small even at exceedingly low oxygen fugacities.

The calculated concentrations of CH<sub>3</sub>Cl (above) are much smaller than those observed in the stratospheric plume from the 18 May 1980 eruption of Mount St Helens<sup>16</sup>. Similarly elevated concentrations of CH<sub>3</sub>Cl at Kilauea are explained<sup>15</sup> by burning organic material, which may also account for the results at Mount St Helens. The burning hypothesis is supported by the following: (1) concentrations of CH<sub>3</sub>Cl were larger in Kilauea volcanic gas samples collected near burning vegetation at the edge of lava flows<sup>15</sup>, (2) CH<sub>3</sub>Cl is known to be produced by burning vegetation<sup>5</sup>, (3) burning vegetation is commonly associated with lava flows or pyroclastic flows, and (4) the calculated equilibrium concentration of halocarbons in the volcanic gases cannot explain the observed concentrations of CH<sub>3</sub>Cl.

Model calculations, which include additional trace metal components, on gases from Merapi<sup>13</sup>, Mount St Helens<sup>21</sup> and Kilauea<sup>22</sup> show that HCl concentrations exceed those of NaCl and KCl in each case. The opposite is true for gases from Mount Etna<sup>21</sup>. Considering that Etna produces a rare alkalic magma, the dominance of NaCl and KCl is an exception among world volcanoes. In gases from most volcanoes HCl and HF are the overwhelmingly dominant species of Cl and F.

The fluxes of HCl and HF from active volcanoes (Table 2) can be estimated using the ratios of HCl and HF to SO<sub>2</sub> in the gas, combined with measurements of the SO<sub>2</sub> flux, determined by correlation spectrometry at both passively degassing and erupting volcanoes<sup>23–24</sup>. Unfortunately, direct measurements of actual eruptions are rare, so estimates of minimum emissions

**Table 3** Estimates of global volcanic and industrial emissions of various gases

| Volcanic emissions                                                   |                                                         |                               |                              |
|----------------------------------------------------------------------|---------------------------------------------------------|-------------------------------|------------------------------|
|                                                                      | SO <sub>2</sub><br>(Tg yr <sup>-1</sup> ) <sup>20</sup> | HCl<br>(Tg yr <sup>-1</sup> ) | HF<br>(Tg yr <sup>-1</sup> ) |
| Global total volcanic emission                                       | 18.7                                                    | 0.4–11                        | 0.06–6                       |
| Degassed into troposphere                                            | 17.0                                                    | 0.3–10                        | 0.05–5                       |
| Explosively erupted into stratosphere                                | 1.66                                                    | 0.1–1                         | 0.005–0.5                    |
| Other sources of HCl and HF                                          |                                                         |                               |                              |
| Sea salt from oceans <sup>30</sup>                                   |                                                         | 300                           | 0.02                         |
| Coal burning <sup>30</sup>                                           |                                                         | 1.8                           | 0.18                         |
| Cl and F as halocarbons                                              |                                                         |                               |                              |
|                                                                      |                                                         | Cl<br>(Tg yr <sup>-1</sup> )  | F<br>(Tg yr <sup>-1</sup> )  |
| Global industrial production of Cl and F as halocarbons <sup>3</sup> |                                                         | 2.28                          | 0.273                        |
| Cl as CH <sub>3</sub> Cl from oceans and burning <sup>3</sup>        |                                                         | 1.4–3.5                       |                              |

of erupted gas from some explosive eruptions (Table 2) have been made by study of melt inclusions<sup>25–26</sup> and by the study of glacial ice cores<sup>27</sup>. Most volcanic eruptions release less Cl and F as HCl and HF than the annual industrial Cl and F production as halocarbons (Table 3). Occasional large explosive events (such as Toba at 75,000 yr BP and Tambora in 1815) may contribute quantities of HCl and HF that are a few orders of magnitude greater than the annual industrial production of Cl and F as halocarbons (Table 3). Although such large eruptions are infrequent, the large Cl loading could have a significant short-range effect on stratospheric chemistry, notably the ozone layer.

We estimate the global volcanic emissions of HCl and HF by using estimates of the volcanic SO<sub>2</sub> flux<sup>20</sup> and data on volcanic gases<sup>15,18,25–27</sup>, which show that the molar ratios of SO<sub>2</sub>/HCl and SO<sub>2</sub>/HF range from nearly 1 to over 100 for volcanism in general. Most convergent plate volcanoes have higher proportions of Cl (ref. 28), so SO<sub>2</sub>/HCl is probably in the range of 1–10. Using 1 and 100 as the range in SO<sub>2</sub>/HCl and SO<sub>2</sub>/HF (a range of 1–10 was for SO<sub>2</sub>/HCl for explosive eruptions at convergent plate boundaries) and an annual SO<sub>2</sub> flux of 18.7 Tg (ref. 20), the annual volcanic emissions of HCl and HF are 0.4–11 Tg and 0.06–6 Tg, respectively. These rates are consistent with other estimates of global volcanic emissions of 0.78–7.8 Tg yr<sup>-1</sup> of HCl and 0.04–0.4 Tg yr<sup>-1</sup> of HF (refs 29–30). They are also consistent with the independent estimates of the liberation of Cl from island arc volcanism over the past 100 million years of 3–30 Tg yr<sup>-1</sup> (ref. 7).

The release of HCl and HF into the troposphere by small explosive eruptions and passively degassing volcanoes (Table 2) represents significant global emissions of 0.3–10 and 0.05–5 Tg yr<sup>-1</sup>, respectively (Table 3). Although sea salt is the main source of tropospheric HCl (refs 3 and 30), volcanoes are (potentially) an important source of tropospheric HF (Table 3), as was previously suggested<sup>30</sup>. These tropospheric emissions may be efficiently removed by precipitation<sup>3</sup>.

To estimate the amounts of HCl and HF that potentially enter the stratosphere, we used the global rates for explosive eruptions. Eruptions with a VEI (volcanic explosivity index) of 4 or higher would penetrate the stratosphere<sup>31</sup>. The average annual production of SO<sub>2</sub> from explosive eruptions with a VEI of 4 or higher has been estimated as 1.66 Tg (ref. 20). Assuming SO<sub>2</sub>/HCl ranges from 1 to 10, these explosive events release 0.1–1 Tg yr<sup>-1</sup> of HCl and 0.005–0.5 Tg yr<sup>-1</sup> of HF (Table 3). These are averages



over many years. A large eruption would cause sudden, sharp loading.

More work is necessary to determine the impact of a large explosive eruption on the ozone layer. As cataclysmic eruptions (VEI of 5 or higher) have column heights<sup>31</sup> of greater than 25 km, the HCl in these plumes should penetrate the maximum concentration of ozone<sup>3</sup> in the stratosphere at 25 km. But the estimates of HCl and HF that potentially enter the stratosphere during explosive volcanic eruptions (Tables 2 and 3) do not account for cooling and condensation, scavenging by ash or other removal processes within water-rich eruption plumes. It is difficult to evaluate what percentage of the gas is removed during the eruption process, although direct measurements on the 1986 explosive eruption plume from Augustine volcano

showed that most of the HCl did not adhere to particles and remained in the gaseous state<sup>24</sup>. A significant fraction of these gases do sometimes enter the stratosphere because the 1982 eruption of El Chichón volcano caused direct stratospheric loading of HCl (ref. 10).

Some<sup>3</sup> suggest that photolysis of anthropogenic halocarbons in the stratosphere is the only major source of atmospheric HF. This paper supports other work<sup>30</sup> that naturally degassing volcanoes also emit significant quantities of HF, some of which is directly injected into the stratosphere. Thus, volcanoes should be regarded as a significant natural source of tropospheric and stratospheric HF.

This work was supported by NSF. We thank A. Bernard for helpful discussions of this work.

Received 16 March; accepted 24 June 1988.

- Molina, M. J. & Rowland, F. S. *Nature* **249**, 810-812 (1974).
- Dickinson, R. E. & Cicerone, R. J. *Nature* **319**, 109-115 (1986).
- WMO/NASA Atmospheric Ozone 1985; Assessment of Our Understanding of the Processes Controlling its Present Distribution and Change, World Meteorological Organization Global Ozone Research and Monitoring Project Report No. 16 1986 (1986).
- Singh, H. B., Salas, L. J. & Stiles, R. E. *J. geophys. Res.* **88**, 3684-3690 (1983).
- Crutzen, P. J., Heidt, L. E., Krasnec, J. P., Pollock, W. H. & Seiler, W. *Nature* **282**, 253-256 (1979).
- Rubey, W. W. *Bull. geol. Soc. Am.* **62**, 1111-1148 (1951).
- Anderson, A. T. *Geol. Soc. Am. Bull.* **85**, 1485-1492 (1974).
- Stolarski, R. S. & Cicerone, R. J. *Can. J. Chem.* **52**, 1610-1615 (1974).
- Johnston, D. A. *Science* **209**, 491-493 (1980).
- Mankin, W. G. & Coffey, M. T. *Science* **226**, 170-172 (1983).
- Woods, D. C., Chuan, R. L. & Rose, W. I. *Science* **230**, 170-172 (1985).
- Giggenbach, W. F. *Bull. volcan.* **39**, 132-145 (1975).
- Symonds, R. B., Rose, W. I., Reed, M. H., Lichte, F. E. & Finnegan, D. L. *Geochim. cosmochim. Acta* **51**, 2083-2101 (1987).
- Stoiber, R. E., Leggett, D. C., Jenkins, T. F., Murrmann, R. P. & Rose, W. I. *Jr Geol. Soc. Am. Bull.* **82**, 2299-2302 (1971).
- Gerlach, T. M. *J. Volcan. geotherm. Res.* **7**, 295-317 (1980).
- Inn, E. C. Y., Vedder, J. F. & Condon, E. P. *Science* **211**, 821-823 (1981).

- Reed, M. H. *Geochim. cosmochim. Acta* **46**, 513-528 (1982).
- Gerlach, T. M. & Casadevall, T. J. *J. Volcan. geotherm. Res.* **28**, 141-160 (1986).
- Rutherford, M. J., Sigurdsson, H., Carey, S. & Davis, A. J. *geophys. Res.* **90**, 2929-2947 (1985).
- Stoiber, R. E., Williams, S. N. & Huebert, B. J. *Volcan. geotherm. Res.* **33**, 1-8 (1987).
- Le Guern, F. thesis, Univ. Paris.
- Naughton, J. J., Lewis, V. A., Hammond, D. & Nishimoto, D. *Geochim. cosmochim. Acta* **38**, 1679-1690 (1974).
- Rose, W. I., Chuan, R. L., Giggenbach, W. F., Kyle, P. R. & Symonds, R. B. *Bull. volcan.* **48**, 181-188 (1986).
- Rose, W. I. *et al. J. geophys. Res.* **93**, 4485-4499 (1988).
- Devine, J. D., Sigurdsson, H. & Davis, A. N. *J. geophys. Res.* **89**, 6309-6325 (1984).
- Sigurdsson, H., Devine, J. D. & Davis, A. N. *Jökull* **35**, 1-8 (1985).
- Hammer, C. U., Clausen, H. B. & Dansgaard, W. *Nature* **288**, 230-235 (1980).
- Anderson, A. T. *Rev. Geophys. Space Phys.* **13**, 37-55 (1975).
- Cadle, R. D. *J. geophys. Res.* **80**, 1650-1652 (1975).
- Cadle, R. D. *Rev. Geophys. Space Phys.* **18**, 746-752 (1980).
- Newhall, C. G. & Self, S. J. *geophys. Res.* **87**, 1231-1238 (1982).
- Millán, M. M., Gallant, A. J., Yong, C. S. & Fauaki, F. *Atmos. Envir.* **19**, 255-263 (1985).
- Greenland, L. P., Rose, W. I. & Stokes, J. B. *Geochim. cosmochim. Acta* **49**, 125-129 (1985).
- Buat-Ménard, P. & Arnold, M. *Geophys. Res. Lett.* **5**, 245-248 (1978).
- Casadevall, T. J. *et al. J. geophys. Res.* **89**, 9633-9641 (1984).
- Phelan, J. M. *et al. Geophys. Res. Lett.* **9**, 1093-1096 (1982).

## Climate forcing implications from Vostok ice-core sulphate data

M. R. Legrand\*, R. J. Delmas\* & R. J. Charlson†

\* Laboratoire de Glaciologie et Géophysique de l'Environnement, BP 96, 38402 St Martin d'Hères cedex, France

† Department of Atmospheric Sciences AK-40, University of Washington, Seattle 98195, USA

The hypothesis that the number concentration of cloud condensation nuclei (mainly the sulphate particles produced by the oxidation of dimethylsulphide emitted from the ocean<sup>1</sup>) influences marine stratus cloud albedo, and hence global climate<sup>2</sup>, is examined using the non-seasalt sulphate ( $\text{nss-SO}_4^{2-}$ ) profile that was recently obtained<sup>3</sup> along the 160 kyr Vostok (Antarctica) ice core. The deduced 20-46% increase in  $\text{nss-SO}_4^{2-}$  content in the Antarctic atmosphere during full glacial (compared with interglacial) conditions is consistent with higher dimethylsulphide emissions from marine biota productivity. Similar spectral features and correlation of  $\text{CO}_2$  and  $\text{nss-SO}_4^{2-}$  suggest a global significance for these changes, and a possible link between  $\text{CO}_2$  and dimethylsulphide emissions. Assuming that the global average number population of cloud condensation nuclei is proportional to  $\text{nss-SO}_4^{2-}$  in the Antarctic atmosphere, a global radiative cooling at the surface of up to 1 K would result, reinforcing the effect of  $\text{CO}_2$  ( $\sim 0.6$  K; ref. 4) and of total insolation changes<sup>5</sup> (0.2 K).

In Fig. 1, the  $\text{nss-SO}_4^{2-}$  profile is shown with deuterium measurements which reflect changes in high latitude surface temperatures over the past 160 kyr<sup>6</sup>. The deuterium curve (Fig. 1a) represents a continuous profile, whereas chemical measurements were performed on 10 cm ice-core sections (spanning between 5 and 10 yr) at 25 m intervals. Non-seasalt sulphate, although mainly of marine biogenic origin, also results from volcanic eruptions that sporadically disturb the background level

of  $\text{H}_2\text{SO}_4$  in the Antarctic<sup>7</sup>. The climatic impact of such short volcanic events is a cooling of the Earth's surface by a few tenths of a degree (see ref. 8 for example) and, although their effect is about one order of magnitude weaker than the expected cloud condensation nuclei (CCN) effect, we first have to ensure that our  $\text{nss-SO}_4^{2-}$  profile is free of significant volcanic perturbations. Additional new measurements have established that the high figures previously found for 225, 250 and 1,550 m depth (shown as broken lines in Fig. 1b) were not related to long-term but to short-term changes, which we attribute to volcanic events. The new  $\text{nss-SO}_4^{2-}$  profile (Fig. 1b) takes into account these modifications. On the other hand, additional measurements performed along stages B, C, D, E and H confirm that the fluctuations of  $\text{nss-SO}_4^{2-}$  previously reported<sup>3</sup> do reflect real fluctuations of the tropospheric background level.

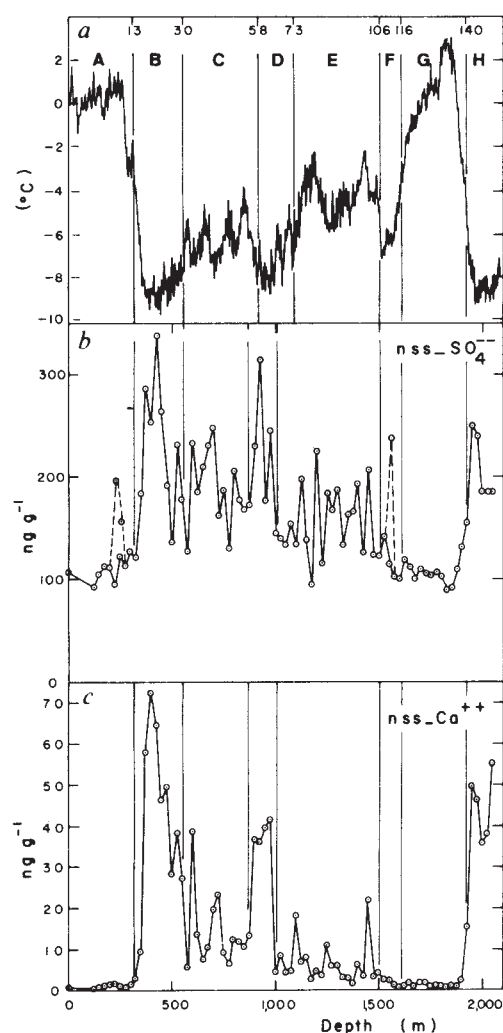
The ionic composition of meltwater samples down the length of this ice core has shown that the main cations associated with the  $\text{nss-SO}_4^{2-}$  are  $\text{H}^+$ ,  $\text{Na}_m^+$  (marine sodium, the reference element for sea salt) and non-seasalt calcium ( $\text{nss-Ca}^{2+}$ , chosen as the

**Table 1** Non-seasalt sulphate for different stages along the Vostok ice core

| Stage depth (m) | $c^*$ ( $\text{ng g}^{-1}$ ) | $R^\dagger$ (%) | $c'^*$ ( $\text{ng g}^{-1}$ ) | $R'^\dagger$ (%) |
|-----------------|------------------------------|-----------------|-------------------------------|------------------|
| A (0-275)       | $108 \pm 10$                 | —               | $108 \pm 10$                  | —                |
| B (350-550)     | $230 \pm 58$                 | 113             | $154 \pm 34$                  | 42               |
| C (575-850)     | $187 \pm 38$                 | 73              | $134 \pm 24$                  | 24               |
| D (875-975)     | $228 \pm 53$                 | 111             | $158 \pm 34$                  | 46               |
| E (1,000-1,475) | $157 \pm 34$                 | 45              | $122 \pm 29$                  | 13               |
| F (1,500-1,600) | $177 \pm 14$                 | 64              | $87 \pm 10$                   | —                |
| G (1,625-1,875) | $105 \pm 10$                 | —               | $106 \pm 5$                   | —                |
| H (1,925-2,050) | $187 \pm 24$                 | 73              | $130 \pm 19$                  | 20               |

\*  $c$  represents concentrations in ice and  $c'$ , concentrations corrected for the accumulation rate effect.

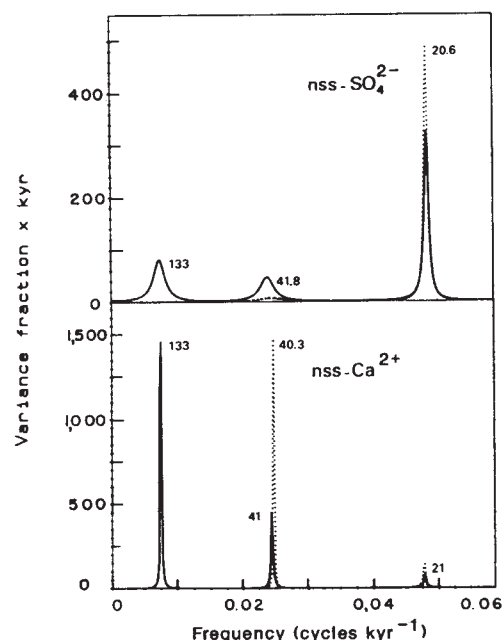
†  $R$  and  $R'$  are changes of  $c$  and  $c'$  with respect to the stage A value.



**Fig. 1** Variations with time of *a*, Vostok isotope temperature record (in °C) (from ref. 6); *b*, non-seasalt sulphate content (in ng g<sup>-1</sup>): dashed line refers to high levels (not confirmed) reported previously<sup>3</sup> (see text); *c*, non-seasalt calcium (the terrestrial contribution, see text and ref. 3). Dating is indicated (in kyr) in the upper part of Fig. 1*a*.

indicator of terrestrial impurities<sup>3</sup>). A detailed discussion of the marine (Na<sub>m</sub><sup>+</sup>) and terrestrial (nss-Ca<sup>2+</sup>, Fig. 1*c*) indicators suggests that the increased concentrations of both elements in ice deposited during glacial ages (stages H and F to B) mainly reflect global changes: an increase in the pole-equator temperature gradient leading both to higher production rates, and to enhanced atmospheric poleward transport<sup>3</sup>. The nss-SO<sub>4</sub><sup>2-</sup> profile shows glacial-interglacial changes from 105–110 ng g<sup>-1</sup> during interglacial stages (A and G) to 190–230 ng g<sup>-1</sup> during very cold stages (B, D and H) (see Table 1).

The higher concentrations of nss-SO<sub>4</sub><sup>2-</sup> found for the ice age are only partly due to the reduced snow accumulation rates during this time. Such an effect is also observed when comparing nss-SO<sub>4</sub><sup>2-</sup> concentrations at several Antarctic locations having similar geographical and meteorological characteristics, but with different snow accumulation (*A*): the lower the accumulation, the higher the nss-SO<sub>4</sub><sup>2-</sup> concentrations. Note that this is also observed for artificial radionuclides<sup>9</sup>, <sup>10</sup>Be (ref. 10), volcanic H<sub>2</sub>SO<sub>4</sub> (ref. 7) and so forth. Using sulphate measurements at different Antarctic locations<sup>11</sup>, we have derived an experimental relation, which was found to be valid both for glacial and interglacial conditions, between *c* (the concentration in ice of nss-SO<sub>4</sub><sup>2-</sup>) and *A*. Then *c'* (the value of *c* corrected for accumula-



**Fig. 2** Variance spectra of the Vostok nss-SO<sub>4</sub><sup>2-</sup> and nss-Ca<sup>2+</sup> content, using corrected values from accumulation rate effect (see text). The normalized variance density is plotted against frequency (in cycles kyr<sup>-1</sup>). Continuous lines refer to the original series and dashed lines to the series after pre-whitening by a first-order difference filter. Variance spectra (maximum entropy method) were obtained as described<sup>5</sup>.

tion changes) can be obtained from *c* using equation (1) and the past accumulation rates reported by Lorius *et al.*<sup>12</sup>.

$$c' = c(0.25 + 0.294 A) \quad (1)$$

where *c* and *c'* are in ng g<sup>-1</sup> and *A* in g cm<sup>-2</sup> yr<sup>-1</sup>. The values of *c'* given in Table 1 are, in our opinion, more representative of actual atmospheric concentrations occurring in the past. In this way we can conclude that atmospheric nss-SO<sub>4</sub><sup>2-</sup> content was enhanced by 20–46% (see Table 1) during full glacial, compared with interglacial, stages.

Two possible explanations exist for the nss-Ca<sup>2+</sup> data (Fig. 1*c*). This element could be injected into the atmosphere as CaSO<sub>4</sub>, which is relatively rare, either on the Earth surface or in aerosols<sup>13</sup>, or as CaCO<sub>3</sub>, which then reacts with H<sub>2</sub>SO<sub>4</sub> and HNO<sub>3</sub> in cloud droplets over the long transport time. Legrand *et al.*<sup>3</sup> noted the absence of carbonates in the ice, and concluded that nss-Ca<sup>2+</sup> is linked with both nss-SO<sub>4</sub><sup>2-</sup> and NO<sub>3</sub><sup>-</sup> and which therefore agrees better with the second assumption. The presence of H<sub>2</sub>SO<sub>4</sub>-containing aerosol particles in the Antarctic atmosphere is related to marine biological activity<sup>11,14</sup>. This emission arises from only a fraction of the total photosynthesis of the oceans and is not limited by seawater SO<sub>4</sub><sup>2-</sup>. Hence this marine sulphur flux would be expected to vary, and could have been substantially larger during the ice age. A recent study of S compounds in Antarctic ice<sup>15</sup> suggests that the Antarctic atmosphere is supplied with H<sub>2</sub>SO<sub>4</sub> (or SO<sub>2</sub> and CH<sub>3</sub>SO<sub>3</sub>H) by long-range transport from lower latitudes far from Antarctica and produced by marine biota. The process by which the fine-particle H<sub>2</sub>SO<sub>4</sub> aerosol (*d* < 0.1 μm) probably reacts with the coarse particles of CaCO<sub>3</sub> and NaCl (*d* > 1 μm) is by in-cloud coalescence of droplets formed from the respective reactants, followed by droplet coalescence and reaction<sup>15</sup>. Although the coalescing process would reduce the number population of CCN, their number population must have been controlled by the H<sub>2</sub>SO<sub>4</sub> component over some fraction of the transport path, as can also be observed today. Even though both the soil dust



and seasalt levels delivered to the ice were enhanced during the glacial age, their mass concentrations were comparable to that of  $\text{nss-SO}_4^{2-}$  at the same time. Because most of the  $\text{nss-SO}_4^{2-}$  particles were smaller ( $d < 0.1 \mu\text{m}$ ) than the soil dust ( $d > 0.5 \mu\text{m}$ , approximately),  $\text{nss-SO}_4^{2-}$  dominated the number population of dust and sea salt by a factor of  $\sim 100$ , at least when close to the  $\text{nss-SO}_4^{2-}$  source regions.

We therefore propose that the enhanced concentration of  $\text{nss-SO}_4^{2-}$  is due to an enhanced oceanic source of dimethylsulphide (DMS), but we must dismiss the possibility that this enhancement arises solely from transport. Changes resulting from current (stage A) conditions, such as enhanced meridional circulation, must have occurred during the entire ice age, as during stage E. Here, assuming no changes in the  $\text{nss-SO}_4^{2-}$  source (some arguments are presented in ref. 17), we can estimate the maximum effect of transport efficiency to be about +13% (see  $R'$  in Table 1). If the transport in E was similar to that during B and D, then the increases of  $\text{nss-SO}_4^{2-}$  deposition (42 and 46% respectively) cannot be explained by transport alone. Moreover, the flux of long-lived (10–100 days) very small particles to the ice is probably best interpreted as a slow leak-out of a large, fast-moving reservoir. Here we expect little or no effect on the flux resulting from enhanced transport, and the concentration of  $\text{nss-SO}_4^{2-}$  in ice and in Antarctic air would be representative of some large fraction of the hemisphere.

Thus, although we cannot yet prove it, we conclude that the increased  $\text{nss-SO}_4^{2-}$  deposition in Antarctic ice during the glacial age indicates that DMS emissions and CCN populations could have been substantially elevated in the planetary boundary layer at lower latitudes. It is impossible to reconstruct accurately the actual number concentrations, because we do not know how the size distribution changed. Further, we do not know the rate at which the  $\text{nss-SO}_4^{2-}$  particles were removed by precipitation or by reaction with  $\text{CaCO}_3$ , and thus reduced the CCN population along the transport path. The lowest decrease in global temperature that could be associated with the increased  $\text{nss-SO}_4^{2-}$  flux would be zero if there were no changes in CCN concentrations and only an increase in  $\text{nss-SO}_4^{2-}$  particle size. The upper limit of this forcing can be estimated by assuming that the increase of  $\text{nss-SO}_4^{2-}$  in the Antarctic atmosphere (in a steady-state transport system) is proportional to an increase in the number concentration of CCN in the mid-latitude marine stratus areas.

A 46% increase in CCN (from the inferred  $\text{nss-SO}_4^{2-}$  atmospheric change, see  $R'$  values in Table 1) at an initial albedo for marine stratus between 0.3 and 0.7 would cause a cloud albedo increase of  $\sim 0.03$ . If marine clouds covered the same area of the Earth as now (31%), the direct radiative effect of this CCN increase produces a cooling of the globe by up to  $\sim 1 \text{ K}$  (this is close to the example of calculated climatic effect in Table 1b of ref. 2). This forcing of climate was maximal for the full glacial conditions of the second part of the last ice age (from 70–15 kyr BP), and remained rather weak (0.15 K) over the beginning of the last glacial age (110–70 kyr BP). In this way, this effect during stages B and D would have been comparable to that resulting from  $\text{CO}_2$  changes ( $\sim 0.6 \text{ K}$ ; ref. 4).

A significant correlation is observed between  $\text{CO}_2$  and  $\text{nss-SO}_4^{2-}$  profiles (80 depth levels,  $r = -0.64$ ) over the last 160 kyr. Besides the large glacial-interglacial contrast, Fig. 1 suggests a cyclic behaviour of these  $\text{nss-SO}_4^{2-}$  changes. Spectral analysis of our  $\text{nss-Ca}^{2+}$  data provides two peaks (Fig. 2) one at 40.3 kyr (close to 41 kyr obliquity component) and another one at 20.6 kyr (close to the precessional component). On the other hand,  $\text{nss-SO}_4^{2-}$  data (all values being corrected for accumulation-rate changes) exhibit a strong peak at 21 kyr and a weak signal at 41.8 kyr (as does  $\text{CO}_2$ , see ref. 4). The quasi-absence of a 41 kyr period (which reflects changes occurring at high latitudes<sup>18</sup> in our  $\text{nss-SO}_4^{2-}$  changes does not contradict the idea that these changes are of global significance. On the other hand, the similarity between the  $\text{CO}_2$  and the  $\text{nss-SO}_4^{2-}$  profile suggests

a link between  $\text{CO}_2$  and DMS production, especially during the second part of the ice age (70–15 kyr), where both the  $\text{CO}_2$  and the  $\text{nss-SO}_4^{2-}$  signal exhibit a strong 21 kyr period. This link could be the change of surface oceanic circulation leading to higher marine productivity. During the first part of the last ice age, we observe a very moderate  $\text{nss-SO}_4^{2-}$  change, whereas  $\text{CO}_2$  is strongly decreased, possibly as a result of a deep circulation change of the ocean<sup>4</sup>. These considerations are not proofs, but they are consistent with the idea that increased DMS emissions may have produced (through cloud albedo) a significant forcing of the climate, with similar amplitude and in phase with  $\text{CO}_2$  changes during the second part of the last ice age. Finally, if this forcing occurred, biospheric and other climatic feedbacks failed to produce an interglacial homeostasis, opposing the idea of a negative feedback<sup>2</sup>.

We thank T. Anderson and S. G. Warren for reviewing the manuscript, J. R. Petit for spectral analysis and J. M. Barnola, C. Genthon, J. Jouzel, C. Lorius and D. Raynaud for discussion. Funding was provided by CNRS (ATP 'Phase atmosphérique des cycles biogéochimiques'). The Vostok ice core was analysed in the framework of a Soviet-French cooperation agreement.

Received 30 March; accepted 22 June 1988.

1. Nguyen, B. C., Gaudry, A., Bonsang, B. & Lambert, G. *Nature* **275**, 637–639 (1978).
2. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
3. Legrand, M., Lorius, C., Barkov, N. I. & Petrov, V. N. *Atmos. Environ.* **22**, 317–331 (1988).
4. Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
5. Genthon, C. *et al.* *Nature* **329**, 414–418 (1987).
6. Jouzel, J. *et al.* *Nature* **329**, 403–408 (1987).
7. Legrand, M. & Delmas, R. J. *Nature* **327**, 671–676 (1987).
8. Kelly, P. M. & Sear, C. B. *Nature* **311**, 740–743 (1984).
9. Lambert, G., Ardouin, B. & Mesbah-Bendezu, A. in *Precipitation Scavenging, Dry Deposition and Resuspension* (eds Pruppacher, B. *et al.*) 1289–1300 (Elsevier, New York, 1983).
10. Raisbeck, G. M. & Yiou, F. *Ann. Glaciol.* **7**, 138–140 (1985).
11. Legrand, M. thesis, Univ. Grenoble (1985).
12. Lorius, C. *et al.* *Nature* **316**, 591–596 (1985).
13. Glaccum, R. A. & Prospero, J. M. *Mar. Geol.* **37**, 295–321 (1980).
14. Cunningham, W. C. & Zoller, W. H. *J. aerosol Sci.* **12**, 367–384 (1981).
15. Saigne, C. & Legrand, M. *Nature* **320**, 240–242 (1987).
16. Andreae, M. O. *et al.* *Science* **232**, 1620–1623 (1986).
17. Sarnthein, M., Winn, K. & Zahn, R. *Abrupt Climatic Changes: Evidence and Implications* (eds Berger, W. H. & Labeyrie, L. D.) 311–337 (Reidel, Dordrecht, 1987).
18. Berger, A. L. & Pestiaux, P. in *Milankovitch and Climate Vol. 1* (eds Berger, A. L. *et al.*) 83–112 (Reidel, Dordrecht, 1984).

## Observation of steric effects in gas-surface scattering

E. W. Kuipers, M. G. Tenner, A. W. Kleyn & S. Stolte\*

FOM-Institute for Atomic and Molecular Physics, Kruislaan 407, 1098SJ Amsterdam, The Netherlands  
\* Physics Department, Catholic University Nijmegen, Toernooiveld, 6525ED Nijmegen, The Netherlands

Molecules adsorbed on a surface are known to show preferential orientations, and in particular, the interaction potential between a linear molecule and a surface depends on the orientation of the molecular axis. But the fact that the molecules eventually adsorbed are orientated with respect to the surface is not evidence that the dynamics of gas-surface interactions is governed by the initial molecular orientation in the gas phase. For example at very low velocities the molecule might achieve its optimum orientation adiabatically during its approach to the surface. Dependence of scattering dynamics on molecular orientation can also be degraded as a result of surface structure and surface vibrations. There have been, experimental studies, however, which suggest that orientational or steric effects could influence gas-surface scattering<sup>1–4</sup>. Thus, the much larger rotational excitation for  $\text{NO}$ - than for  $\text{N}_2$ -scattering from  $\text{Ag}(111)$  indicates that large anisotropies can occur in a potential for gas-surface dynamics (G. O. Sitz *et al.*

manuscript in preparation). The double rotational rainbow observed in the rotational state distribution for NO/Ag(111) (refs 3 and 4) can also be interpreted as a manifestation of this anisotropy<sup>5-9</sup>. These studies all indicate that steric effects could be important and here we report on scattering of orientated NO from Ag(111) to provide direct experimental evidence for the importance of such effects in gas-surface interactions.

In our experiments we use a pulsed supersonic molecular beam. Because of the low rotational temperatures (4 K) in these beams, nearly all molecules will be in the lowest rotational level of the  $\pi_{1/2}$  electronic ground state of NO. At electric fields strong enough for  $\Lambda$ -decoupling ( $E > 12 \text{ kV cm}^{-1}$ ), the orbital motion of the unpaired electron spin induces an essentially linear Stark effect for all  $M_J$  components of the  $J = 1/2$  rotational ground state<sup>10</sup>. As a result of this Stark effect, the electric hexapole focusing technique<sup>11</sup> can be used to produce a beam of state-selected and orientated NO molecules in an electric guiding field normal to the Ag(111) surface<sup>12,13</sup>. The guiding field is generated by placing a 3-mm diameter rod parallel to the face of the crystal at a high voltage. The orientational distribution of the molecules with respect to the electric field in the  $^2\Pi_{1/2}(J = 1/2, \Omega = 1/2, M_J = 1/2)$  state selected is shown in Fig. 1. It can be seen from this diagram that by choosing a positive rod voltage, the N-end of the molecule can be preferentially aimed at the surface. A negative rod voltage causes an orientational distribution with the O-end chiefly towards the surface. For negative, zero and positive rod voltages, we have measured, the angular distribution of the in-plane reflected molecules at an incident translational energy of  $E_i = 200 \text{ meV}$ , at an angle of incidence of  $\theta_i = 37^\circ$ , measured with respect to the surface normal, and at a surface temperature of  $T_s = 600 \text{ K}$ . In-plane scattering was chosen because it probes essentially only those molecules that are reflected directly from the crystal and have not desorbed after trapping<sup>14</sup>. In Fig. 2 we plot  $[I_+(\theta_r) - I_0(\theta_r)]/I_0(\theta_r)$  and  $[I_-(\theta_r) - I_0(\theta_r)]/I_0(\theta_r)$  as a function of  $\theta_r$ , the reflection angle measured with respect to the surface normal. Here  $I_+(\theta_r)$  is the scattered intensity measured at  $\theta_r$  with a positive rod voltage (thus with the N-end preferentially pointing towards the surface before the collision). For  $I_-(\theta_r)$  and  $I_0(\theta_r)$  respectively, the rod voltage was negative (with the O-end preferentially towards the surface) and zero (random orientation). From this figure we see that the rod voltage influences the direct scattering of NO from Ag(111). For small scattering angles, we observe that  $I_-(\theta_r) < I_0(\theta_r) < I_+(\theta_r)$ , whereas for large scattering angles the reverse is seen. This means that for the N-end collision, the preferred scattering angle is smaller, whereas for the O-end collision, it is larger than for a random-orientation collision. Assuming the angular distribution for direct inelastic scattering for random orientation to be  $I_0(\theta_r) = \cos^{10}(\theta_r - \theta_i - 5^\circ)$  (refs 14 and 15), we obtain the angular distributions for positive and negative rod voltages shown in Fig. 3. These angular distributions have been calculated by increasing or decreasing  $I_0(\theta_r)$  by the percentages shown in Fig. 2. The shift between the angular distributions can be explained in the following way.

Near-specular scattering indicates that the surface used in our experiments is flat and structureless<sup>14</sup>. Thus the molecular momentum (velocity component) parallel to the surface is expected to remain unchanged by the collision. This means that rotational excitation can only occur at the expense of the normal component of the incident velocity and that the resulting loss of this normal momentum component implies an inclination of  $\theta_r$  towards the surface. Such changes in inclination have been measured for NO-Ag(111) with state-specific detection<sup>4,15</sup>. Apparently our observations show that O-end collisions produce more rotational excitation than N-end collisions. Numerical scattering calculations on highly anisotropic interaction potentials<sup>5-9</sup>, could essentially reproduce the rotational excitation spectra observed<sup>4,16</sup>. The highest rotational excitation or rotational rainbow was found to occur for an incoming orientation of the molecular axis of  $45^\circ$  (ref. 5) and the O-end pointed

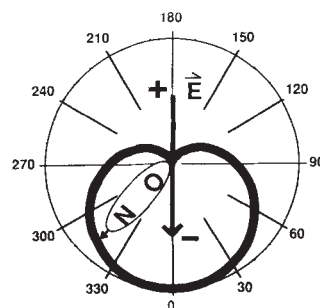


Fig. 1 Polar plot of the probability distribution for the angle between the molecular axis of NO( $^2\Pi_{1/2}(J = 1/2, \Omega = 1/2, M_J = 1/2)$ ) with respect to an external E-field high enough for  $\Lambda$ -decoupling ( $E > 12 \text{ kV cm}^{-1}$ ).

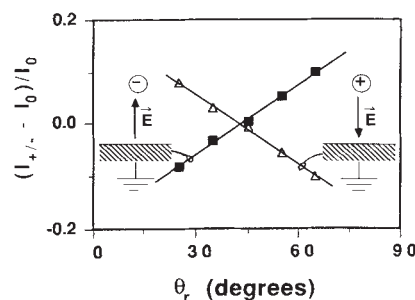


Fig. 2 Results of the measurements of NO( $^2\Pi_{1/2}(J = 1/2, \Omega = 1/2, M_J = 1/2)$ ) direct scattering from a Ag(111) surface, without guiding field and with an electric guiding field of  $15 \text{ kV cm}^{-1}$  for  $\theta_i = 37^\circ$ ,  $E_i = 200 \text{ meV}$  and  $T_s = 600 \text{ K}$ . The relative change between the intensity scattered under the reflection angle  $\theta_r$  with an electric guiding field,  $I_{+/-}(\theta_r)$ , and without an electric guiding field,  $I_0(\theta_r)$ , is plotted. The squares are the measurements with negative rod voltage, E pointing away from the surface, and the triangles are the measurements for a positive rod voltage, E pointing towards the surface. The lines through the points are drawn for guidance.

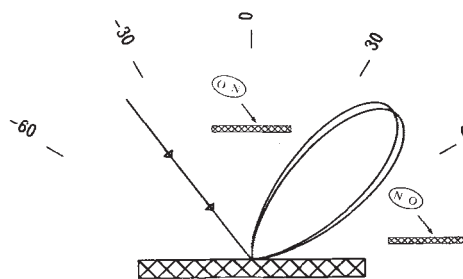


Fig. 3 Angular distributions,  $I_+(\theta_r)$  and  $I_-(\theta_r)$ , for direct scattering of NO from Ag(111), presented in a polar plot for the two different orientation distributions of the NO molecular axis with respect to the Ag(111) surface ( $\theta_i = 37^\circ$ ,  $E_i = 200 \text{ meV}$ ,  $T_s = 600 \text{ K}$ ). The angular distributions,  $I_+(\theta_r)$  and  $I_-(\theta_r)$ , are deduced from the measurements shown in Fig. 2, assuming  $I_0(\theta_r) = \cos^{10}(\theta_r - \theta_i - 5^\circ)$  (refs 14, 15). Plotted in the inserts are the initial average orientations for the two measurements, which are deduced from Fig. 1 and the direction of the electric field.

towards the surface<sup>6</sup>. The rotational excitation with the same  $45^\circ$  orientation, but with N in front, was calculated to be much less. This agrees with our results shown in Fig. 3. Although more complex models, including energy transfer to phonons and a structured surface, may change the detailed shape of the angular distributions, we believe that the physical picture presented here will remain unchanged.



Steric effects have been found to govern the mechanism of purely gas phase reactions<sup>17,18</sup>. In the present study we provide experimental proof that effects of molecular orientation are also important in gas-surface collisions. Orientated molecule scattering from surfaces could provide a method for determining molecule surface anisotropic potentials.

We thank Dr J. T. Hougen, Dr A. E. de Vries and Professor F. W. Saris for critical reading of the manuscript. This work is part of the research program of the Foundation for Fundamental Research of Matter and was supported by the Dutch Organization for Advancement of Research.

Received 25 April; accepted 7 June 1988.

- Novakoski, L. V. & McClelland, G. M. *Phys. Rev. Lett.* **59**, 1259-1262 (1987).
- Sitz, G. O., Kummel, A. C. & Zare, R. N. *J. Vac. Sci. Technol.* **A5**, 513-517 (1987).
- Kleyn, A. W., Luntz, A. C. & Auerbach, D. *J. Phys. Rev. Lett.* **47**, 1169-1172 (1981).
- Kleyn, A. W., Luntz, A. C. & Auerbach, D. *J. Surf. Sci.* **117**, 33-41 (1982).
- Voges, H. & Schinke, R. *Chem. Phys. Lett.* **100**, 245-250 (1983).
- Muhlhausen, C. W., Williams, L. R. & Tully, J. C. *J. chem. Phys.* **83**, 2594-2606 (1985).
- Smedley, J. E., Corey, G. C. & Alexander, M. H. *J. chem. Phys.* **87**, 3218-3231 (1987).
- Tanaka, S. & Sugano, S. *Surf. Sci.* **136**, 488-502 (1984).
- Brenig, W., Kasai, H. & Müller, H. *Surf. Sci.* **161**, 608-626 (1985).
- Stolte, S. thesis, Univ. Nijmegen (1972).
- Kramer, K. H. & Bernstein, R. B. *J. chem. Phys.* **42**, 767-770 (1965).
- Ende, D. van den & Stolte, S. *Chem. Phys. Lett.* **76**, 13-15 (1980).
- Ende, D. van den & Stolte, S. *Chem. Phys.* **89**, 121-139 (1984).
- Kuipers, E. W., Tenner, M. G., Spruit, M. E. M. & Kleyn, A. W. *Surf. Sci.* **189/190**, 669-674 (1987).
- Retner, C. T. *et al. J. Vac. Sci. Technol.* **A5**, 508-512 (1987).
- Luntz, A. C., Kleyn, A. W. & Auerbach, D. *J. Phys. Rev.* **B25**, 4273-4275 (1982).
- Special issue on dynamical stereochemistry, *J. Phys. Chem.* **91**, (1987).
- Levine, R. D. & Bernstein, R. B. *Molecular Reaction Dynamics and Chemical Reactivity* (Oxford University Press, New York, 1987).

## Experimental verification of episodic acidification of freshwaters by sea salts

R. F. Wright\*, S. A. Norton†, D. F. Brakke‡  
& T. Frogner\*

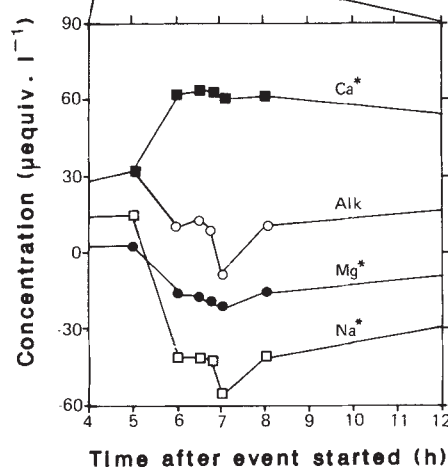
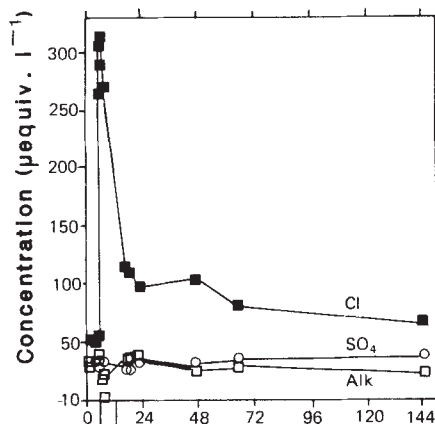
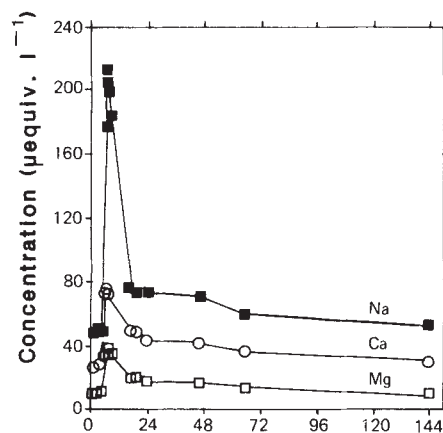
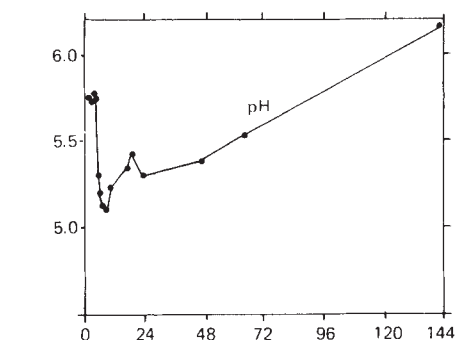
\* Norwegian Institute for Water Research, Box 33, 0313 Oslo, Norway

† Department of Geological Sciences, University of Maine, Orono, Maine 04469, USA

‡ Institute for Watershed Studies, Western Washington University, Bellingham, Washington 98225, USA

Short-term acidification of streams and lakes can cause fish kills and other adverse biological effects<sup>1,2</sup>. Such acid episodes may be due to acid deposition or the result of natural processes. The 'seasalt effect' is a natural process in which episodic input of seasalt-rich precipitation to acid soils can cause acidification of runoff<sup>3</sup>. Several cases of acid episodes observed in coastal regions of Norway<sup>4</sup> and eastern United States<sup>5</sup> have been ascribed to the seasalt effect. To test experimentally this effect we dosed a small pristine catchment in western Norway with dilute sea water. Runoff chemistry responded immediately; pH dropped from 6.1 to 5.1, alkalinity from 20 to -2  $\mu\text{equiv. l}^{-1}$ , and labile monomeric aluminium increased from 15 to 95  $\mu\text{g l}^{-1}$ . This experiment verifies that high inputs of sea salts can cause natural acidification episodes due to cation exchange in small catchments with acidic soils.

The seasalt effect can be ascribed to cation-exchange processes in soils<sup>6,7</sup>, and has been demonstrated in laboratory experiments with soil columns<sup>8,9</sup> and plot experiments<sup>10</sup>. A fraction of the incoming Na and Mg in sea salt is exchanged for other cations in the soil. In acid soil a significant fraction of the exchanged cations will be the acid cations H<sup>+</sup> and Al, and thus the pH and alkalinity of the soil solution and surface waters will be affected. The role of sea salts in acid episodes in streams and lakes has been inferred from field measurements<sup>4,5</sup>. Empirical evidence for the seasalt effect comes from concentrations of Na



**Table 1** Chemical composition of natural precipitation, applied water and runoff from control and seasalt-treated catchments at Sogndal

|                  | Control catchments<br>1984-87 |        | Seasalt experiment July 1987<br>Runoff |        |      |       | MAGIC simulation<br>Runoff |      |
|------------------|-------------------------------|--------|----------------------------------------|--------|------|-------|----------------------------|------|
|                  | Precip.                       | Runoff | Added                                  | Before | Peak | After | Before                     | Peak |
| H <sup>+</sup>   | 16                            | 2      | 1                                      | 1      | 8    | 4     | 1                          | 8    |
| pH               | 4.8                           | 5.7    | 6.0                                    | 6.1    | 5.1  | 5.4   | 6.2                        | 5.1  |
| Ca               | 6                             | 18     | 36                                     | 26     | 72   | 48    | 25                         | 130  |
| Mg               | 8                             | 11     | 122                                    | 10     | 38   | 20    | 10                         | 54   |
| Na               | 38                            | 46     | 520                                    | 48     | 204  | 73    | 48                         | 113  |
| K                | 6                             | 3      | 11                                     | 1      | 2    | 1     | 3                          | 6    |
| Al               | —                             | 0      | 0                                      | 0      | 8    | 1     | 0                          | 7    |
| NH <sub>4</sub>  | 9                             | 2      | —                                      | —      | —    | —     | 5                          | 5    |
| SO <sub>4</sub>  | 23                            | 24     | 80                                     | 23     | 21   | 16    | 22                         | 26   |
| Cl               | 41                            | 43     | 605                                    | 42     | 302  | 99    | 42                         | 279  |
| NO <sub>3</sub>  | 8                             | 1      | 0                                      | 0      | 3    | 0     | 2                          | 2    |
| HCO <sub>3</sub> | —                             | 8      | 9                                      | 30     | 0    | 22    | 25                         | 2    |
| Σ <sup>+</sup>   | 83                            | 82     | 690                                    | 86     | 332  | 147   | 92                         | 323  |
| Σ <sup>-</sup>   | 81                            | 76     | 694                                    | 95     | 326  | 137   | 91                         | 329  |
| alk.             | -5                            | 12     | 4                                      | 20     | -2   | 27    | 22                         | -19  |
| TOC              | —                             | 1.2    | 1.2                                    | 1.7    | 1.5  | 1.0   | —                          | —    |
| SiO <sub>2</sub> | —                             | 1.1    | 1.1                                    | 3.6    | 2.2  | 2.8   | —                          | —    |
| Ali              | —                             | 10     | 0                                      | 15     | 95   | 30    | 0                          | 76   |

Data for precipitation and runoff at the control catchments are from Wright *et al.*<sup>12</sup>. Average yearly precipitation and runoff at Sogndal are ~1,000 and 900 mm, respectively. For the seasalt experiment total dose on 21 July from 16:00–21:30 was 15 mm and runoff on 21–22 July totalled 2 mm. Runoff data are for samples collected 21 July 13:15 (before), 21 July 22:15 (peak) and 22 July 10:00 (after). Also shown are results of simulation using the MAGIC model<sup>15,16</sup>. Inorganic Al (Ali) was measured as labile monomeric Al<sup>18</sup>. Concentrations of ionic Al species were calculated using the procedure of Schecher and Driscoll<sup>19</sup>; values here are the sum of cationic species. Alkalinity (alk.) is calculated as the difference in sum of base cations minus sum of strong acid anions. Dashes indicate no data. Units:  $\mu\text{equiv. l}^{-1}$ ; total organic carbon mg C  $\text{l}^{-1}$ ; SiO<sub>2</sub> mg  $\text{l}^{-1}$ ; Ali  $\mu\text{g l}^{-1}$ .

relative to Cl; when the equivalent ratio of Na/Cl is less than that of sea water (that is, the concentration of non-marine Na is negative), it can be assumed that ion exchange has occurred. The absence of negative values for non-marine Na, however, does not preclude significant acidification by the seasalt effect. Because seasalt-driven acid episodes in streams are short-lived, they are seldom measured by normal sampling routines. In lakes the effect is moderated by the residence time of water. To verify, clarify and quantify the processes involved in a seasalt-driven acidification we experimentally added sea salts to a small pristine catchment with acid soils.

We conducted our seasalt experiment at a 0.38-ha headwater catchment located near Sogndal, western Norway, in conjunction with the acid-addition experiments of the Reversing Acidification In Norway (RAIN) project<sup>11,12</sup>. The catchment is 900 m above sea level and is characterized by alpine vegetation and thin, acidic (pH 4.6–5.4), poorly developed soils (lithic haplumbrept) on siliceous gneissic bedrock<sup>13</sup>. The area presently receives nearly unpolluted precipitation with a volume-weighted pH of 4.8. Surface waters are typical of regions sensitive to acid deposition with pH 5.5–6.5, alkalinity 10–25  $\mu\text{equiv. l}^{-1}$ , and sum of non-marine base cations 25–40  $\mu\text{equiv. l}^{-1}$  (Table 1)<sup>11,12</sup>. Natural precipitation events with high levels of sea salts occur 2–3 times each year with loading of Cl > 0.6 mequiv.  $\text{m}^{-2}$  and concentrations of Cl > 150  $\mu\text{equiv. l}^{-1}$ .

We produced a seasalt episode by spraying the catchment with 11 mm of water containing 820  $\mu\text{equiv. l}^{-1}$  Cl. A commer-

cial sprinkling system was used. We mixed sea water with lake water from the adjacent control catchment SOG1 at a ratio 1:700. The catchment was pre-wetted with 2 mm of lake water before salt treatment and washed down with 2 mm following the treatment. Application was at 2 mm  $\text{h}^{-1}$ . A total of 15 mm with average Cl concentration of 605  $\mu\text{equiv. l}^{-1}$  was added (Table 1). This is the same ionic strength and water application rate as used in the acid-addition experiments conducted on nearby catchments<sup>11,12</sup>.

The experiment was conducted after a two-week dry period. Soils were dry and of the 15 mm water added, only about 2 mm left the catchment in runoff. Discharge began ~4 h after onset of watering and ceased 24 h later.

Seasalt addition caused prompt and dramatic changes in runoff chemistry (Table 1, Fig. 1). Chloride concentrations increased from 42 to 302  $\mu\text{equiv. l}^{-1}$  within two hours of the onset of runoff. The increase in Cl was accompanied by increases in the base cations Ca, Mg, Na and K, increases in the acid cations H<sup>+</sup> and Al, a decrease in alkalinity from 20 to -20  $\mu\text{equiv. l}^{-1}$ , and a decrease in pH from 6.1 to 5.1. Twelve hours after watering had ceased runoff chemistry had recovered to a large extent, although full recovery of alkalinity took about three weeks.

A similar volume of water added to the adjacent catchment caused only a minor increase in Cl concentration in runoff (from 28 to 33  $\mu\text{equiv. l}^{-1}$ ). This increase is probably due to evapotranspiration and dry deposit of seasalts during the previous two-week dry period.

The nearly complete retention of SO<sub>4</sub> in the salt experiment is in agreement with results from the sulphuric acid addition experiments at the adjacent catchments<sup>11,12</sup> and can be explained by sulphate adsorption.

Runoff before treatment contained 42  $\mu\text{equiv. l}^{-1}$  Cl and reached 302  $\mu\text{equiv. l}^{-1}$  Cl during the period immediately following seasalt application (Table 1). Under the assumption that Cl is inert in the soil, this runoff can be envisaged as a mixture of 15 mm applied water with 605  $\mu\text{equiv. l}^{-1}$  Cl and 18 mm soil solution containing 42  $\mu\text{equiv. l}^{-1}$  Cl. Runoff 12 h later had 100  $\mu\text{equiv. l}^{-1}$  Cl and indicates mixing with an additional

◀ **Fig. 1** The pH and concentrations of major cations, anions and alkalinity in runoff at a small headwater catchment near Sogndal, western Norway, before, during and after experimental addition of diluted sea water. Also shown are concentrations of excess base cations (Na\*, Ca\*, Mg\*) (non-marine fraction calculated from Cl concentration and chemical composition of sea water) during the runoff peak. Alkalinity is calculated as the difference between sum of charges of base cations minus sum of charges of strong acid anions. Negative concentrations of excess base cations indicate that a fraction of these cations in seasalt input has been retained in the soil relative to Cl.



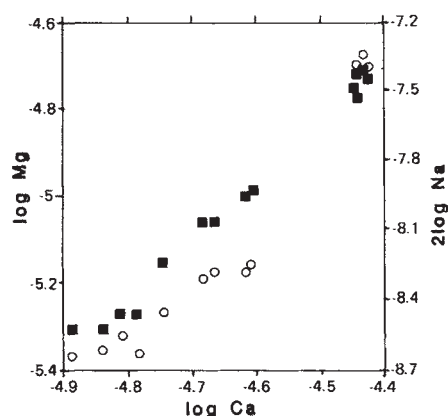


Fig. 2 Plot of log Mg (squares, scale at left) and 2 log Na (circles, scale at right) against log Ca concentrations ( $\text{mol l}^{-1}$ ) in runoff during the seasalt addition experiment. The close linear relationships indicate that variations are largely explained by cation exchange in the soils.

110 mm soil solution. Together these data indicate a total of 130 mm soil water before treatment, an amount consistent with average soil depth of 30 cm, porosity of 50% and bulk density (dry) of  $1,350 \text{ kg m}^{-2}$  (ref. 13).

The changes in Cl concentrations suggest that two types of soil water interact with the applied sea salts to give runoff. The first soil water comes as quickflow and may be present in soil macropores, whereas the second soil water comes as baseflow and may reflect capillary water in the soil. The distinction between soil macropore and micropore water at the Sogndal site comes from studies of soil solution and runoff at the acid-treated catchments, and is supported by tracer studies with LiBr (R. Parnell, personal communication).

The changes in concentrations of cations in runoff are due primarily to cation exchange in the soil. The mix of cations in runoff is not explained by simple dilution of applied sea salt with soil water, by the change in ionic strength, or by the formation of ion pairs. The concentration of excess Na (non-marine fraction) in runoff dropped from  $15 \mu\text{equiv. l}^{-1}$  before treatment to  $55 \mu\text{equiv. l}^{-1}$  and excess Mg declined from 5 to  $20 \mu\text{equiv. l}^{-1}$ , whereas excess Ca increased from 25 to  $65 \mu\text{equiv. l}^{-1}$  (Fig. 1). The relationship between the desorption of Ca and adsorption of Mg is nearly perfectly explained by cation exchange (Fig. 2). The relationship between Ca and Na approximately follows that expected, with cation exchange assuming the Gaines-Thomas form<sup>14</sup> (Fig. 2), although here other factors also appear to be of importance. The decrease in alkalinity in runoff suggests an increase in Al in soil solution also due to cation exchange<sup>7</sup>.

Chemical weathering of soil minerals cannot account for the short-term changes in runoff chemistry. Cation exchange responds quickly to changes in soil solution chemistry, whereas chemical weathering is a decidedly slower process and probably more important for soil water with long retention time. The absence of significant change in  $\text{SiO}_2$  concentrations during the experiment also indicates that chemical weathering plays a minor role over the short term.

We used MAGIC, a model for soil and water acidification<sup>15,16</sup>, to simulate the seasalt addition experiment. MAGIC used soil-chemical processes including cation exchange, chemical weathering,  $\text{CO}_2$  dissolution, sulfate adsorption and dissolution of Al taken at the catchment scale to account for observed responses in soil and water chemistry to changes in the chemical composition of atmospheric precipitation. The model was previously calibrated to the control catchments at Sogndal and used to evaluate the acid-addition experiments in the adjacent catchments<sup>17</sup>. The model correctly predicts direction and approximate degree of response to seasalt addition including the exchange

of Na and Mg for  $\text{H}^+$ , Ca and Al, and decreases in alkalinity and pH (Table 1). At peak Cl concentrations of  $302 \mu\text{equiv. l}^{-1}$ , however, the model predicts Na concentrations of only  $113 \mu\text{equiv. l}^{-1}$ , a concentration considerably lower than the observed  $204 \mu\text{equiv. l}^{-1}$ . A fraction of the treated water may reach the stream as runoff without participating in soil ion-exchange reactions, or the macropore water may be in contact with soil which has different cation selectivity coefficients or fractions of exchangeable cations than the bulk soil. The MAGIC model is apparently too coarse in both time and space to simulate quantitatively such short-term episodes, and indeed it was not designed to do so<sup>15,16</sup>.

Our experiment simulated a natural precipitation event with high concentrations of sea salts such as occurs in coastal regions. Concentrations of sea salts in the applied water were only 2–3 times higher than maximum levels observed between 1983 and 1987 in weekly bulk precipitation at the Sogndal site. Cation exchange in the acid soils was the most important process leading to the short-term acidification of runoff. The response to seasalt addition was rapid, and the pH of runoff decreased as ion-exchange reactions occurred in the soil. The acidification was short-lived, although full recovery took from days to weeks. This experiment indicates that neutral sea salts can produce significant short-term loss of alkalinity in sensitive freshwaters. The seasalt effect cannot, however, account for long-term acidification of surface waters, and is of importance only on an episodic basis in coastal regions with thin, acidic soils.

This work was supported in part by the Royal Norwegian Council for Scientific and Industrial Research.

Received 12 January; accepted 21 March 1988.

- Leivestad, H. & Muniz, I. P. *Nature* **259**, 391–392 (1979).
- Harvey, H. H. & Whelpdale, D. M. *Wat. Air & Soil Pollut.* **30**, 579–586 (1986).
- Norton, S. A. et al. in *GEOMON International Workshop on Geochemistry and Monitoring in Representative Basins*, 148–150 (Geological Survey, Prague, 1987).
- Skartveit, A. *Nordic Hydrol.* **12**, 65–80 (1981).
- Kahl, J. S. et al. *Water Resource Baseline Data and Assessment of Impacts from Acidic Precipitation, Acadic National Park, Maine* (National Park Service Technical Report 16, Washington, 1985).
- Wiklander, L. *Geoderma* **14**, 93–105 (1975).
- Reuss, J. O. *J. environ. Qual.* **12**, 591–595 (1983).
- Johnson, D. W. et al. *Soil Sci. Soc. Am. J.* **50**, 776–783 (1986).
- Njøs, A. *Solute and Water Transport Through Soils with Special Reference to Raw Humus Layers* (Internal Report 39/78, SNSF-project, NISK, 1432 Ås, Norway, 1978).
- Christophersen, N. et al. *Nordic Hydrol.* **13**, 105–114 (1982).
- Wright, R. F. et al. *Wat. Air & Soil Pollut.* **30**, 47–64 (1986).
- Wright, R. F. et al. *Nature* (in the press).
- Lotse, E. & Ottabong, E. *Physicochemical properties of soils at Risdalsheia and Sogndal: RAIN project* (Acid Rain Res. Rep. 8/1985, Norwegian Institute for Water Research, Oslo, 1985).
- Gaines, G. L. & Thomas, H. C. *J. chem. Phys.* **21**, 714–718 (1953).
- Cosby, B. J. et al. *Wat. Resour. Res.* **21**, 51–63 (1985).
- Cosby, B. J. et al. *Wat. Resour. Res.* **21**, 1591–1601 (1985).
- Wright, R. F. & Cosby, B. J. *Atmos. Envir.* **21**, 727–730 (1987).
- Røgeberg, E. J. S. & Henriksen, A. *Vatten* **41**, 48–53 (1985).
- Schecher, W. D. & Driscoll, C. T. *Wat. Resour. Res.* **23**, 525–534 (1987).

## Exceptional preservation of fossils in an Upper Proterozoic shale

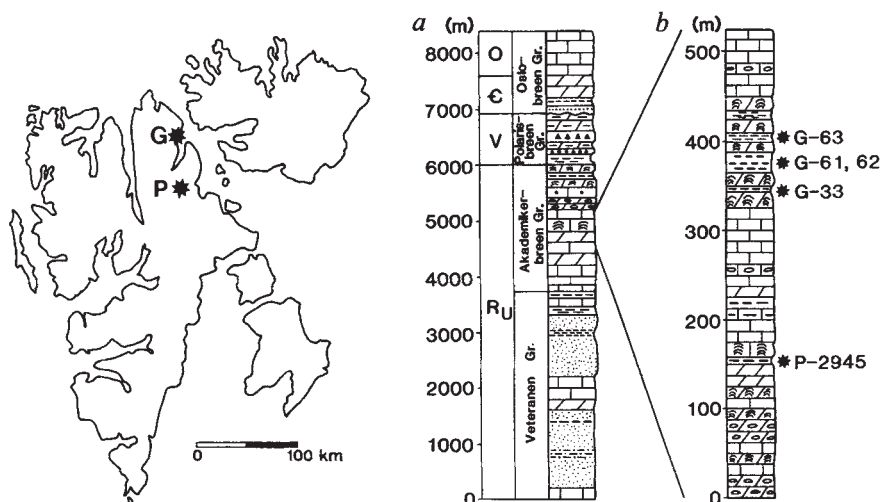
N. J. Butterfield, A. H. Knoll & K. Swett\*

Botanical Museum, Harvard University, Cambridge, Massachusetts 02138, USA

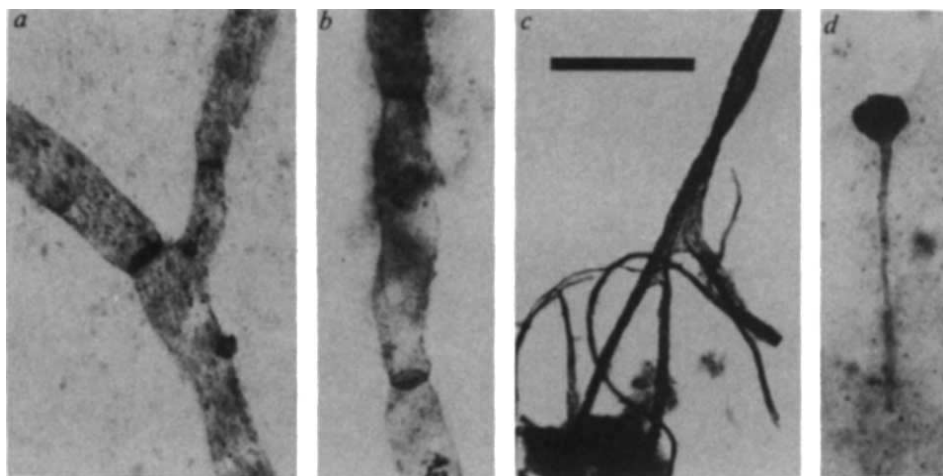
\* Department of Geology, University of Iowa, Iowa City, Iowa 52242, USA

Late Proterozoic organisms must have been diverse and widely distributed, but in general their fossil record is both taxonomically and environmentally limited. Exceptional preservation of Proterozoic fossils is not unknown, but it is usually associated with silicified carbonates from restricted peritidal or playa lake environments<sup>1,2</sup>. We report here an exceptionally well preserved and distinctive assemblage of Late Proterozoic fossils from subtidal

**Fig. 1** Map of Spitsbergen showing locations of fossiliferous shales in coastal exposures at Geerabukta (G) and in nunatak walls along Polarisbreen glacier (P). Also shown are geological columns illustrating the general stratigraphy of the Upper Proterozoic Akademikerbreen Group and the stratigraphy of the Svanbergfjellet Formation in a section measured at Geerabukta. Stratigraphic positions of fossiliferous shales are indicated by asterisks followed by sample numbers.



**Fig. 2** *a, b*, Multicellular algae comparable to some extant cladophoralean green algal species (bar in *c* = 60  $\mu\text{m}$  for *a*, 100  $\mu\text{m}$  for *b*). *c*, Robust, branching structure comparable to the prostrate holdfasts of certain extant chaetophoralean green algae; for example, *Stigeoclonium pusillum* (bar = 150  $\mu\text{m}$ ). *d*, Spheroidal vesicle with filamentous extension, tentatively identified as the germinating zoospore of a filamentous protist (bar in *c* = 60  $\mu\text{m}$ ). *a, b* and *d* are from bedding-parallel thin sections; *c* is from an acid maceration.



marine shales. In addition to the sphaeromorphic acritarchs and cyanobacterial sheaths routinely preserved in Proterozoic rocks, this assemblage includes multicellular algae ('seaweeds'), a diverse assortment of morphologically complex protistan vesicles, and probable heterotrophic bacteria. Thus, it provides one of the clearest and most taxonomically varied views of Proterozoic life yet reported.

The fossils occur in the Svanbergfjellet Formation, a 550–650-m succession of carbonates and subordinate shales exposed in the glaciated mountains and coastal cliffs of north-eastern Spitsbergen<sup>3</sup> (Fig. 1). Fossiliferous horizons occur within a distinctive 100–200-m sequence of massive stromatolitic biostromes interbedded with subequal thicknesses of shale and, lower in the formation, within a 5-m shale bed immediately beneath a laterally extensive stromatolitic bench. The fossiliferous shales are green to black, entirely siliciclastic, and extremely fine grained, with most mineral grains significantly finer than 1  $\mu\text{m}$ . Sub-millimetric lamination indicates that they were deposited in quiet subtidal waters protected from strong wave activity. Thin clay laminae provided rapid and efficient barriers to continued physical and biological degradation. Regular lamination also facilitates study by means of bedding-parallel thin sections<sup>4</sup>, which retain not only delicate and otherwise unrecoverable fossils, but also biologically significant details of fossil orientation and bedding-plane distribution.

Although the age of the Svanbergfjellet fossils is poorly constrained by radiometric data, a Late Riphean age of approximately 700–800 Myr BP is indicated on several grounds. The fossiliferous sediments sit some 2,200 m below lowermost Cambrian strata and 1,100–1,300 m beneath the base of Vendian

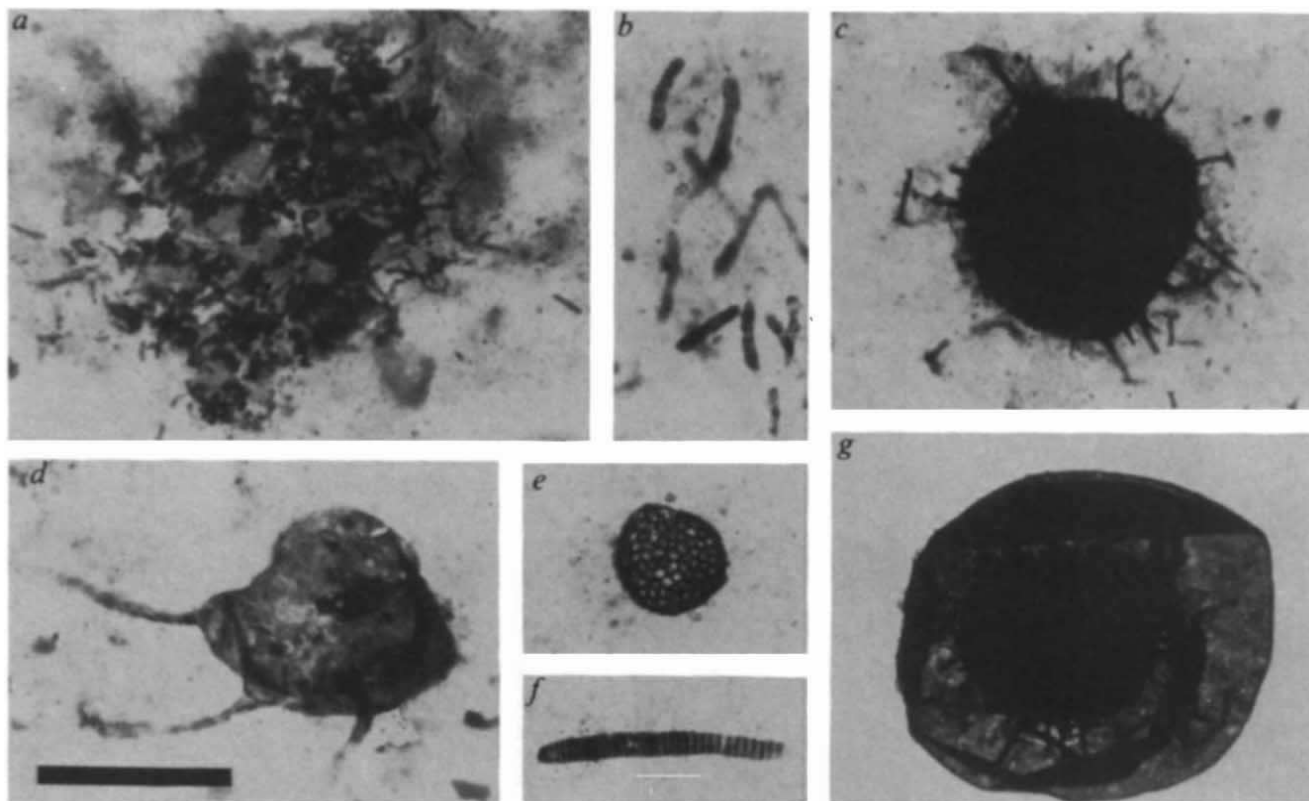
glacial beds; both stromatolites and acritarchs in this and associated formations are distinctly Late (but not latest) Riphean in aspect<sup>5–7</sup>; and carbon and strontium isotopic ratios for Svanbergfjellet carbonates are similar to those reported for Upper Riphean rocks elsewhere, but distinctly different from reported Vendian or Palaeozoic values (refs 8, 9 and unpublished data).

Metaphytes ('seaweeds') are a conspicuous component of the Svanbergfjellet shale benthos. One population exhibits extensive, branching filamentous thalli composed of large (6–50  $\times$  50–800  $\mu\text{m}$ ) cylindrical cells. Adjacent cells are attached via distinctively thickened septal plates (Fig. 2*a, b*). Branching is intercalary, with branch insertion near the apices of axial cells (Fig. 2*a*). It is impossible to ascertain the complete height of the algae from which these fossils were derived, but preserved fragments are up to 1 cm long. This population has close morphological analogues among extant cladophoralean green algae (Ulvophyceae), including species that today form 'algal carpets' below wave base in tropical reef areas<sup>10</sup>.

More enigmatic are robust, complex branched structures 10–60  $\mu\text{m}$  in diameter and up to 1 mm long found in the lower shale horizon (Fig. 2*c*). These show no obvious cellularity, consisting instead of robust fibrils that branch repeatedly. Among living algae, the rhizoids or hold-fasts of certain of the heterotrichous Chaetophorales (Cholophyceae, *sensu stricto*) provide a close structural analogue.

Solitary filaments ( $\approx 5 \times 100 \mu\text{m}$ ) that terminate in bulbous ( $\approx 35 \mu\text{m}$ ) swellings occur sporadically in thin sections (Fig. 2*d*); these may be related developmentally, or at least be comparable functionally, to large (30–110  $\mu\text{m}$ ) vesicles bearing 2–7 tubular extensions (Fig. 3*d*). Given their marked asymmetry





**Fig. 3** *a*, Cluster of probable heterotrophic bacterial rods. *b*, Higher power photomicrograph showing the morphology of individual rods. *c*, *g*, Large process-bearing acritarch, which in *g* is preserved within a larger unornamented vesicle. *d*, Vesicle with three asymmetrically placed filamentous extensions, comparable to the germinating zoospores of some extant protists, for example, *Vaucheria sessilis*. *e*, Ornamented acritarch, comparable in general morphology to Paleozoic fossils assigned to the genus *Acrum*. *f*, Well preserved oscillatory cyanobacterial trichome. The fossils in *a-f* are from bedding-parallel thin sections; *g* is from an acid maceration. Bar in *c* = 50  $\mu\text{m}$  for *a*, *c*, *d* and *e*; = 20  $\mu\text{m}$  for *b*; = 110  $\mu\text{m}$  for *f*; and = 170  $\mu\text{m}$  for *g*.

and pliant walls, these structures resemble, and may well represent, germinating zoospores of filamentous protists—modern analogues include the xanthophycean alga *Vaucheria*.

Oscillatorian and chroococcalean cyanobacteria in the shales compare closely with those reported from silicified Proterozoic carbonates. Isolated, short trichomes (Fig. 3*f*) may be allochthonous, but dense tangles of filaments that cover bedding planes indicate that *in situ* benthic microbial mats were common in muddy environments, just as they were in coeval carbonates.

Geochemical signatures indicate that heterotrophic bacteria were present in Proterozoic communities, but their remains have seldom been observed directly<sup>11</sup>. Svanbergfjellet shales contain populations of rod-shaped and filamentous fossils that simply on the basis of size (<2  $\mu\text{m}$ ) might be considered as possible heterotrophs. More persuasive is the preserved distribution (inferred behaviour) of these populations within individual laminae. For example, small (1–1.5  $\times$  5–10  $\mu\text{m}$ ) rods occur repeatedly as dense colonies within local patches of organic detritus (Fig. 3*a, b*), with individuals often aligned toward or along discrete organic particles. Other possible heterotrophic bacteria include clusters and chains of smaller rods (0.4  $\times$  0.7  $\mu\text{m}$ ), occasionally branching filaments that resemble extant actinomycetes.

Superimposed on these benthic associations are a wide variety of presumed allochthonous remains. Megascopic forms include *Chuar*, *Tawuia*, and thin-walled spheroidal vesicles 200–1,500  $\mu\text{m}$  in diameter. Smaller unornamented acritarchs occur as isolated individuals, small clusters, or large populations of medially split spheroids that cover entire bedding planes, suggesting the episodic occurrence of plankton blooms or mass encystments. More distinctive are the at least eight taxa of

morphologically complex acritarchs. In addition to the Late Riphean genus *Trachyhystrichosphaera*, the Svanbergfjellet assemblage includes forms comparable to the predominantly Palaeozoic genera *Acrum* (Fig. 3*d*) and *Micrhystridium*, large process-bearing vesicles resembling late Vendian acritarchs assigned to *Comasphaeridium* and *Baltisphaeridium*, and several previously unreported forms. Particularly striking are large (150–250  $\mu\text{m}$ ), opaque, bristle-covered vesicles that bear long (up to 60  $\mu\text{m}$ ), terminally flared processes (Fig. 3*c*). Occasionally, these are preserved within still larger (300  $\mu\text{m}$ ) unornamented vesicles, strongly supporting their interpretation as encystment structures (Fig. 3*g*). Their phylogenetic relationships remain unclear.

In concert with equally distinctive, but specifically different acritarch assemblages in latest Proterozoic shales and cherts from China<sup>12,13</sup> and Australia (W. Zang and M. R. Walter, personal communication), Svanbergfjellet acritarchs suggest the potential for a much enhanced Upper Proterozoic microfossil zonation. Coupled with emerging data on isotopic chemostratigraphy<sup>14</sup>, magnetostratigraphy<sup>15</sup>, and sequence stratigraphy<sup>16</sup>, complex acritarchs hold promise for a major improvement in our ability to resolve Late Proterozoic time.

Palaeobiologically, the Svanbergfjellet biota provide evidence for a substantially pre-Ediacaran radiation of at least two classes of metaphytic green algae, as well as morphologically complex unicellular protists. At the same time, the fine lamination preserved in the shales indicates that metazoans larger than 50–100  $\mu\text{m}$  in diameter were not present in at least the local benthos. Perhaps most important, the Svanbergfjellet fossils demonstrate that a much wider sampling of Proterozoic life is available than has hitherto been appreciated.

This research was supported in part by grants from the NSF and NASA, an NSERC pre-doctoral fellowship to N.J.B., and a John Simon Guggenheim Foundation fellowship to A.H.K.

Received 17 March; accepted 16 June 1988.

1. Knoll, A. H. *Phil. Trans. R. Soc.* **311B**, 111–122 (1985).
2. Southgate, P. N. *Geology* **14**, 683–686 (1986).
3. Wilson, C. B. *Geol. Mag.* **98**, 89–116 (1961).
4. Horodyski, R. J. & Bloesser, B. *Science* **199**, 682–684 (1978).
5. Jankauskas, T. V. in *Stratotype of the Riphean: Paleontology and Paleomagnetism* (ed. Keller, B. M.) 84–120 (USSR Acad. Sci. Geol. Inst. Nauka, Moscow, 1982).
6. Vidal, G. & Knoll, A. H. *Geol. Soc. Am. Mem.* **161**, 265–277 (1983).
7. Raaben, M. E. & Zabrodin, V. E. *Tr. Geol. Inst. Akad. Nauk S.S.S.R.* **217**, 1–130 (1972).
8. Knoll, A. H. *et al. Nature* **321**, 832–838 (1986).
9. Veizer, J. *et al. Geochim. cosmochim. Acta* **47**, 295–302 (1983).
10. Round, F. E. *The Ecology of Algae* (Cambridge Univ. Press, Cambridge, 1981).
11. Lanier, W. P. J. *sed. Petrol.* **58**, 89–99 (1988).
12. Awaramik, S. M. *et al. Nature* **315**, 655–659.
13. Yin, L. *Palaeont. Cathayana* **2**, 229–249 (1984).
14. Magaritz, M. Holser, W. T. & Kirschvink, J. L. *Nature* **320**, 258–259 (1986).
15. Kirschvink, J. L. *Geol. Mag.* **115**, 139–150 (1978).
16. Christie-Blick, N., Grotzinger, J. P. & van der Borch, C. C. *Geology* **16**, 100–104 (1988).

## Evidence from flipper structure for a single origin of pinnipeds

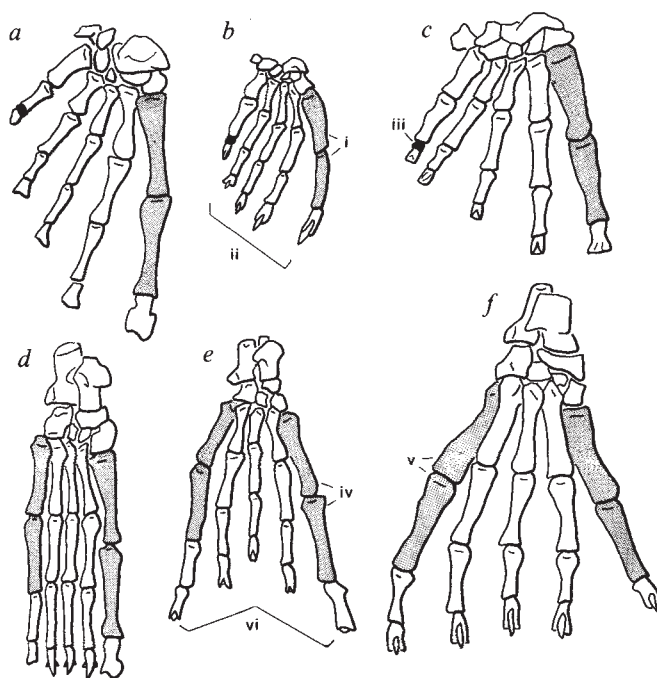
André R. Wyss

Department of Vertebrate Paleontology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA

The mutual affinities of pinnipeds (the 'fin-footed' members of the placental mammal order Carnivora) have been the subject of controversy for more than a century<sup>1,2</sup>. Disagreement centres over whether the origin of the group from a terrestrial ancestry involved a single or multiple marine invasion(s). Particularly since influential papers of the late 1950s and '60s (refs 3, 4), the dual origin of pinnipeds—sea lions (otariids) plus walruses (odobenids) from bear-like forms, and 'true' seals (phocids) from weasel- or otter-like forms—has become the overwhelmingly dominant view in the morphological systematic literature<sup>5–8</sup>, a point reflected by its acceptance in several general texts<sup>9–11</sup>. Anatomical evidence from a survey of appendicular osteology presented here contradicts this notion and suggests radical alteration of currently accepted hypotheses of the relationship of the major pinniped lineages. This evidence strongly suggests a single (monophyletic) origin for the group, a finding in agreement with certain biomolecular results<sup>12–15</sup>.

Somewhat anomalously, flippers (structures well marked enough to lend pinnipeds their name) have scarcely been examined in a phylogenetic context. This lack of attention stems from a common misperception that organs of propulsion and other specializations clearly associated with an aquatic habitat are too evolutionarily labile to be accorded serious phylogenetic consideration. For this reason, numerous fine-scaled morphological resemblances of pinnipeds have gone either ignored or unrecognized. Results presented here, both in terms of the degree of detailed morphological correspondence and functional considerations, assuage concerns that pinniped resemblances simply represent convergently acquired adaptations to an aquatic existence. Instead, these resemblances are more judiciously interpreted as bespeaking a common heritage.

Several highly peculiar features of the pinniped locomotory apparatus that have previously gone unreported in this connection, lend support to the monophyly view. Conspicuous resemblances between pinniped limbs reside in the osseous structure of the flippers themselves. Evident from the comparative views of the manus in representative members of the three living pinniped families (Fig. 1a, b, c) are: the elongation and stout development of digit I, the progressive decrease in emphasis of digits I–V, and the marked reduction of intermediate



**Fig. 1** Comparison of the right fore and hind feet of representative members of the three extant groups of pinnipeds: (a, d), *Eumetopias jubatus* (Otariidae); (b, e), *Monachus tropicalis* (Phocidae); (c, f), *Odobenus rosmarus* (Odobenidae). The figure shows several potential pinniped synapomorphies: (i) marked elongation and enlargement of both major elements of first digit of manus (shaded); (ii) progressive decrease in emphasis of digits I–V on manus; (iii) strong reduction of fifth intermediate phalanx of manus (shown in black); (iv) strong development and elongation of first metatarsal and proximal phalanx (shaded); (v) well developed and elongated major elements of digit V on pes; (vi) weak development of digit III. Emphasis of metatarsal V in e and f supports the recently proposed odobenid-phocid alliance<sup>22</sup>.

phalanx V. Additional similarities not discernible from the figure include: phalangeal flattening and non-trochleated, hinge-like interphalangeal articulations, and reduction of the median keel on the disto-plantar surface of the metacarpal heads and their associated pits on the proximal phalanges.

Turning to the foot, here again occur several distinctive pinniped attributes (Fig. 1d, e, f): strong development and elongation of metatarsal I and the first proximal phalanx, well developed and elongate major elements of digit V, weak digit III, and a similar pattern of metapodial-phalangeal and interphalangeal articulations discussed above for the hand.

These advanced pedal features in pinnipeds also occur in all extinct lineages for which adequate material is known<sup>16,17</sup>. Among living forms, the only exceptions to the above-noted distribution of features occur among some phocine phocids: these incongruities, however, clearly represent secondary modifications which cannot be regarded as primitive phocid features<sup>18</sup>.

By contrast to the conditions obtaining in pinnipeds just outlined, in terrestrial carnivorans digits I of the manus and I and V of the pes are the most reduced of their respective limbs (the central digits being those most emphasized), the fifth intermediate phalanx of the manus is of appreciable size, phalanges are rounded in cross-section and their articulations trochleated, and metapodial heads bear strong keels.

Additional uniquely derived and equally highly visible attributes of the proximal appendicular skeleton of pinnipeds include: well developed greater and lesser trochanters, a strong deltoid-pectoral crest, and marked shortening of the humerus (also present in cetaceans); short, antero-posteriorly flattened femur with fovea for teres femoralis ligament lacking or greatly



reduced; olecranon process of ulna spatulate; and dorso-ventrally broad distal half of the radius.

The suite of modifications of the pinniped appendicular skeleton enumerated above does not resemble that seen in the limbs of other aquatic mammals, nor is this constellation of features duplicated elsewhere among vertebrates, Recent or extinct. Thus, although flippered appendages may well represent obligate structural requirements of an aquatic tetrapod (consider their independent appearance in sirenians, cetaceans, chelonians, and so on) appendages of a very particular design can not be dismissed so easily. The mere association of a given evolutionary innovation with a particular lifestyle does perforce diminish its phylogenetic relevance. Aquatic specializations are not necessarily 'superficial' and close agreement in subtle anatomical detail may reflect genealogy irrespective of its correlation with environment. Taken together, the detailed resemblances of the pinniped flipper and proximal limb argue strongly for an exclusive common ancestry for the group.

A functional perspective lends further credence to the single-origin interpretation. Proposals invoking independent adaptation mediated by functionally imposed requirements of an aquatic existence to account for the remarkably close overall correspondence in pinniped limb osteology face contradiction from what is known about pinniped locomotor behaviour. It is well established that odobenids and phocids on the one hand, and otariids on the other, use their limbs in the water in an utterly different manner, with thrust being produced in the former pair of taxa by alternating strokes of the hindlimbs<sup>19</sup>, and in otariids by the forelimbs<sup>20</sup>. Thus, apart from descent through common ancestry, the development of structures so anatomically identical yet functionally completely dissimilar is a matter not easily dealt with. Significant also is the fact that these limb characters concord with the distribution of numerous pinniped cranial features (including: details of ossicular morphology, the lack of a pit for insertion of the tensor tympani muscle, round window greatly enlarged over oval window, basal whorl of scala tympani enlarged, and maxillary contribution to anterior rim and wall of orbit<sup>21,22</sup>) equally suggestive of monophyly.

Of further importance to the acceptance of a single pinniped origin is the paucity of evidence supporting the close relationship of one pinniped subgroup with any particular terrestrial lineage. The need to hypothesize widespread convergence between independent pinniped clades (as acceptance of the diphyly notion requires)<sup>4-8</sup> is obviated by recent studies showing the morphological basis of the proposed alliance of phocids and mustelids to be far less secure than is generally assumed<sup>22,23</sup>.

Although pinnipeds exhibit numerous features conforming with an aquatic habitat, the detailed components of these features are of such unique morphology that there is little cause to regard them as having arisen convergently. Special resemblances of limb structure as well as a host of other derived morphological features, appear to compellingly corroborate the hypothesis of pinniped monophyly.

I thank R. H. Tedford, M. C. McKenna, M. J. Novacek and M. A. Norell for discussion and comments. Research was supported by a faculty fellowship (Columbia University) and by the Frick Laboratory Endowment Fund (American Museum of Natural History).

Received 21 March; accepted 14 June 1988.

1. Mivart, G. *Proc. zool. Soc. Lond.* XX, 484-500 (1885).
2. Kellogg, R. *Bull. Dept. Geol. Univ. Calif.* 13, 23-132 (1922).
3. Savage, R. J. G. *Proc. zool. Soc. Lond.* 129, 151-244 (1957).
4. McLaren, I. A. *Syst. Zool.* 9, 18-28 (1960).
5. Tedford, R. H. *Syst. Zool.* 25, 363-374 (1976).
6. Muizon, C. de. *Geobios Mém. Spéc.* 6, 259-277 (1982).
7. Muizon, C. de, *Ann. S. Afr. Mus.* 89, 175-213 (1982).
8. Barnes, L. G., Domning, D. P. & Ray, C. E. *Mar. Mamm. Sci.* 1, 15-53 (1985).
9. King, J. E. *Seals of the World* Second edition (Cornell University Press, Ithaca, 1983).
10. *The encyclopedia of mammals* (ed. MacDonald, D.) (Facts on File, New York, 1984).
11. Savage, R. J. G. & Long, M. R. *Mammal evolution: an illustrated guide* (British Museum, London, 1986).
12. Fay, F. H., Rausch, V. R. & Felitz, E. T. *Can. J. Zool.* 45, 773-778 (1967).

13. Sarich, V. M. *Syst. Zool.* 28, 286-295 (1969).
14. Jong, W. W. de, in *Macromolecular Sequences in Systematics and Evolutionary Biology* (ed. Goodman, M.) 75-114 (Plenum, New York, 1982).
15. Arnason, U. & Widegren, B. *Molec. Biol. Evol.* 3, 356-365 (1986).
16. Mitchell, E. *Univ. Calif. Publ. geol. Sci.* 61, 1-105 (1966).
17. Repenning, C. A. & Tedford, R. H. *U.S. Geol. Sur. Prof. Paper* 992, 1-93 (1977).
18. Wyss, A. R. *Am. Mus. nat. Hist. Novitates* (in the press).
19. Gordon, K. R. *J. Zool.* 195, 349-367 (1981).
20. English, A. W. *J. Zool.* 178, 341-364 (1976).
21. Repenning, C. A. in *Functional Anatomy of Marine Mammals* Vol. 1 (ed. Harrison, R. J.) 307-331 (Academic, New York, 1972).
22. Wyss, A. R. *Am. Mus. nat. Hist. Novitates* 2871, 1-31 (1987).
23. Wiig, O. *Zool. Scripta* 12, 225-227 (1983).

## Evidence for modality-specific meaning systems in the brain

Rosaleen A. McCarthy\* & E. K. Warrington

National Hospital for Nervous Diseases, Queen Square, London WC1N 3BG, UK

Patients with cerebral lesions offer a unique opportunity to investigate the organization of meaning systems in the brain. Clinical neurologists have long been aware that knowledge of particular classes or categories of information may be selectively impaired in some cases and selectively spared in others. For example, knowledge of letters, colours, objects, or people may be lost as a consequence of damage to the left hemisphere of the brain<sup>1-4</sup>. Recently there has been quantitative evidence for even more specific impairment and preservation of particular classes of knowledge<sup>5-8</sup>. More recently the evidence for knowledge of living things as compared with inanimate objects is particularly striking<sup>9-12</sup>. Such observations have suggested that our semantic knowledge base is categorical in its organization. In this preliminary report, we describe a patient whose semantic knowledge deficit was not only category specific, but also modality specific. Although his knowledge of the visual world was almost entirely normal, his knowledge of living things (but not objects!) was gravely impaired when assessed in the verbal domain. These findings call into question the widely accepted view that the brain has a single all-purpose meaning store<sup>13</sup>.

Category-specific effects are most plausibly considered as arising from a deficit within the central meaning system itself. The alternative possibilities, namely failure at an earlier stage of processing, or disruption of the links between earlier (pre-semantic) processes and the meaning system can be discounted. For a category-specific deficit to arise in the first place it is necessary that the information should have already been categorized along a semantic dimension. Evidence for a degradation of categories of semantic knowledge, rather than failure to access an intact set of meaning representations, is based on two further converging sources of evidence. Firstly, patients may be highly consistent over testing sessions with regard to the specific items that they can or cannot comprehend. Such a pattern of response would be unlikely to arise simply as a product of 'random noise' in the meaning system, or in access to it. This evidence is more plausibly considered as reflecting the loss of particular items of information<sup>11</sup>. Secondly, the loss of certain items is not absolute. The errors made by such patients indicate partial knowledge of the target, so that broad category information is preserved (that is, a type of bird, vegetable, object and so on), but at the same time more detailed knowledge is lost (such as colour, size, type of material). Thus a patient would know that a canary was a bird, but not know that it was yellow<sup>5</sup>. This invariant hierarchy of partial knowledge is difficult to account for in terms of a deficit at a presemantic level of processing, but is entirely compatible with a degradation of the meaning representations themselves.

\* Present address: Department of Experimental Psychology, Downing Street, Cambridge CB2 3EB, UK.

The question remains as to whether these findings can be explained in terms of a disorder to an all-purpose meaning system, or whether there are a number of dissociable meaning systems corresponding to different sensory modalities (auditory or visual for example). The majority of cases with category-specific disorders have shown some impairment in comprehending both the spoken word and in demonstrating knowledge of the same class of material when it is presented in pictorial form<sup>13</sup>. However, there may be a less than perfect concordance between the two sensory modalities, both in terms of the severity of deficit and the specific items that are affected<sup>11</sup>. The best-documented modality-specific disorder is the syndrome of visual associative agnosia, in which the patient is unable to assign meaning to objects or pictures presented visually, but may be able to perform normally when asked to define their appropriate verbal labels. For example, the patient may be unable to recognize a hammer, but have no difficulty in giving an adequate definition of the spoken word 'hammer'<sup>14</sup>. This single dissociation, although it is suggestive of there being modality-specific meaning systems, could be incorporated into a model postulating a single all-purpose meaning system, by postulating a disconnection syndrome<sup>11</sup>. However, a modality-specific impairment in which there was not only clear evidence of category specificity, but which also showed the hallmarks of a degradation of the meaning representations, would be strong evidence for multiple modality-specific meaning systems.

The patient we have investigated (T.O.B.) was a 63-year-old senior civil servant who complained of progressive deterioration in his use of language and comprehension of the spoken word. There was no other abnormality on neurological examination. PET (positron emission tomography) scanning demonstrated a relatively well circumscribed abnormality in the left temporal lobe (P. J. Tyrell, D. Perani and M. N. Rossor, personal communication). He spoke freely and fluently, using a somewhat repetitive and limited vocabulary. He frequently resorted to stereotyped phrases in his attempts to circumvent his obvious word-finding difficulties. Very occasional minor errors of syntax were also noted. He was tested on the Wechsler Adult Intelligence Scale and obtained an average verbal IQ of 95 and a high average performance IQ of 116. The only feature of note was his total failure to define a number of relatively common words which must have once been well within his vocabulary. This deficit accounted for the relative depression of his verbal IQ. His reading and writing were also somewhat impaired. His performance was also very weak on formal tests of naming: for example, he failed to score at all on a Graded Difficulty Naming Test. He had equal difficulty in matching a spoken name to a picture. Clinical observation suggested that his difficulty was not simply one of word retrieval, but that in addition certain spoken names appeared to be almost totally meaningless to him. This contrasted with an excellent ability to derive meaning from a picture. For example, when asked to define the word 'rhinoceros', he responded, "Animal, can't give you any functions." But when shown a picture of a rhinoceros he responded, "Enormous, weighs over one ton, lives in Africa". Asked to define 'dolphin', he replied, "A fish or a bird"; but when shown a picture he responded, "Dolphin lives in water ... they are trained to jump up and come out ... In America during the war years they started to get this particular animal to go through to look into ships." It was further observed that he appeared to have much greater difficulty in comprehending the names of animate things as compared with inanimate objects. For example, when asked to define 'pig', he responded, "Animal", and when asked to define 'lighthouse', he replied, "Round the coast, built up, tall building, lights revolve to warn ships." His definition of the word 'wheelbarrow' was, "The item we have here in the gardens for carrying bits and pieces; wheelbarrows are used by people doing maintenance here on your buildings. They can put their cement and all sorts of things in it to carry it around." By contrast, for peacock he replied, "Common name,

**Table 1** Picture and word identification†

|               | Pictures | Words |
|---------------|----------|-------|
| Living things | 94%      | 33%   |
| Objects       | 98%      | 89%   |

† % Correct.

n = 48.

can't place it." These observations were formally tested using a matched set of items divided into two categories, namely living things (36 animals and 12 plants) and objects<sup>11</sup>. T.O.B. was asked either to define a spoken word, or to explain what was represented in a picture. His responses were recorded on video and scored by independent raters using a lenient criterion that only the core concept need be conveyed. The percentage correct in each condition for the two categories is shown in Table 1. He was clearly impaired in only one of the four conditions, namely the definition of the spoken names of living things.

Although they are among the first words in a child's vocabulary, the names of animals tend to be of lower frequency in adult language than the names of objects. T.O.B. was given a second, purely verbal version of the definition task, using new stimuli in which the material from animate and inanimate categories was precisely matched for their frequency in the language. He was tested on two separate occasions to establish the degree of consistency in his responses. The percentage correct for each category for each condition is shown in Table 2. These findings replicate those from the previous test in showing a selective impairment in T.O.B.'s knowledge of the spoken names of animals. Furthermore, when re-tested he showed a highly consistent pattern of performance, failing the same items on each occasion. Indeed, his responses were only discrepant on 5 of the 56 object stimuli and 3 of the 56 animal stimuli.

In summary, we report evidence that T.O.B. has a category-specific impairment that affects in particular his knowledge of spoken names of living things, and that this impairment is consistent across time. These effects are prominent in the verbal domain, but at the same time his visual knowledge of the same items appears relatively unaffected. The view that the brain has a single 'all-purpose' meaning store is inadequate as an account of this pattern of results<sup>13</sup>. Firstly, the fact that it is one semantic category rather than another that is affected places this deficit within the domain of meaning. Secondly, the fact that T.O.B.'s impairment is consistent from one occasion to another clearly indicates that he shows degradation within the meaning representations themselves, rather than a failure of access to them. Therefore the observation that T.O.B.'s degradation deficit is confined to one modality and to one category within that modality (namely, living things) not only refutes the notion of an 'all-purpose' meaning store, but also provides positive evidence for multiple meaning representations.

T.O.B.'s knowledge of the visual world appears to be relatively intact, whereas his verbal knowledge is gravely impaired within one semantic category. We have argued that this pattern of performance indicates that there are dissociable modality-specific meaning systems in the brain. It has previously been suggested that the differential salience of information from different sensory and motor channels in the acquisition of knowledge is at the basis of the categorization organization of meaning<sup>10</sup>. The modality-specific organization of meaning can plausibly be viewed as having a similar developmental basis to that of

**Table 2** Word definitions†

|             | Animals | Objects |
|-------------|---------|---------|
| 1st attempt | 30%     | 64%     |
| 2nd attempt | 36%     | 70%     |

† % Correct.

n = 56.



category specificity. We would further speculate that this basic categorical organization of meaning provides a blueprint, which is to some extent duplicated in functionally independent modality-specific meaning systems. Indeed, the basic facts of phylogeny and ontogeny make such an organization eminently reasonable, if not inevitable.

We are grateful to Drs J. E. Rees and M. N. Rossor for permission to investigate T.O.B., a patient under their care. We wish to acknowledge the MRC Cyclotron Unit, Hammersmith Hospital and the staff of the Psychology Department, National Hospital. R.A.M.'s research was funded by Cambridge University.

Received 20 April; accepted 15 June 1988.

1. Lissauer, H. *Cognitive Neuropsychol.* **5**, 153-192 (1988).
2. Nielson, J. M. in *Memory and Amnesia* (San Lucas, Los Angeles, 1958).
3. Konorski, J. in *Integrative activity of the brain: An interdisciplinary approach* (University of Chicago Press, 1967).
4. Coughlan, A. K. & Warrington, E. K. *Brain* **101**, 163-185 (1978).
5. Warrington, E. K. *Quart. J. exp. Psychol.* **27**, 635-657 (1975).
6. Goodglass, H. & Budin, C. *Neuropsychologia* **26**, 67-78 (1988).
7. McCarthy, R. A. & Warrington, E. K. *Neuropsychologia* **23**, 709-727 (1985).
8. Hart, J., Berndt, R. S. & Caramazza, A. *Nature* **316**, 439-440 (1985).
9. Warrington, E. K. & McCarthy, R. A. *Brain* **106**, 859-878 (1983).
10. Warrington, E. K. & McCarthy, R. A. *Brain* **110**, 1273-1296 (1987).
11. Warrington, E. K. & Shallice, T. *Brain* **107**, 829-853 (1984).
12. Sartori, G. & Job, R. *Cognitive Neuropsychol.* **5**, 105-132 (1988).
13. Riddoch, M. J., Humphreys, G. W., Coltheart, M. & Funnell, E. *Cognitive Neuropsychol.* **5**, 3-25 (1988).
14. McCarthy, R. A. & Warrington, E. K. *J. Neurol. Neurosurg. Psychiatr.* **49**, 1233-1240 (1986).

## Inhibitory tagging system facilitates visual search

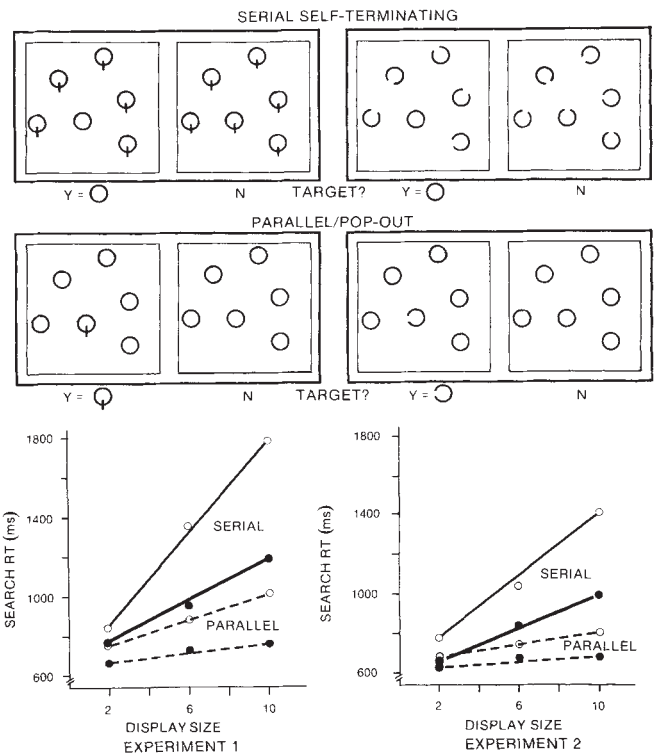
Raymond Klein

Department of Psychology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Two visuospatial phenomena, serial search and inhibition of return, have recently gained the attention of scientists from such diverse disciplines as neuroscience, artificial intelligence and cognitive psychology<sup>1-5</sup>. A linear increase in search latency with increasing display size has been assumed to reflect serial focused attention to each item in the display. A delay in the detection of a signal in a previously attended location has been assumed to reflect an inhibitory process that may be used to prevent attention from returning to the same stimulus. The following study of human performance supports these assumptions and, by demonstrating that inhibition of return operates in serial search, presumably to improve search efficiency, provides a functional link between these two phenomena.

Events in the visual field often attract attention as indicated by a nearly immediate facilitation for the processing of nearby stimuli<sup>6-10</sup>. In contrast, Posner and Cohen<sup>1</sup> discovered a delayed inhibitory effect which begins about 200 ms after a visual event so long as attention is not maintained at the location of that event. They called this 'inhibition of return' to indicate its proposed function, which is to decrease the likelihood that attention would return to previously attended locations. Inhibition of return is coded in environmental rather than retinal coordinates and can last for at least 1.3 s (refs 1, 11, 12). These properties, which rule out masking explanations, suggest that inhibition of return may function to increase the efficiency of visual search<sup>1,7,11</sup>.

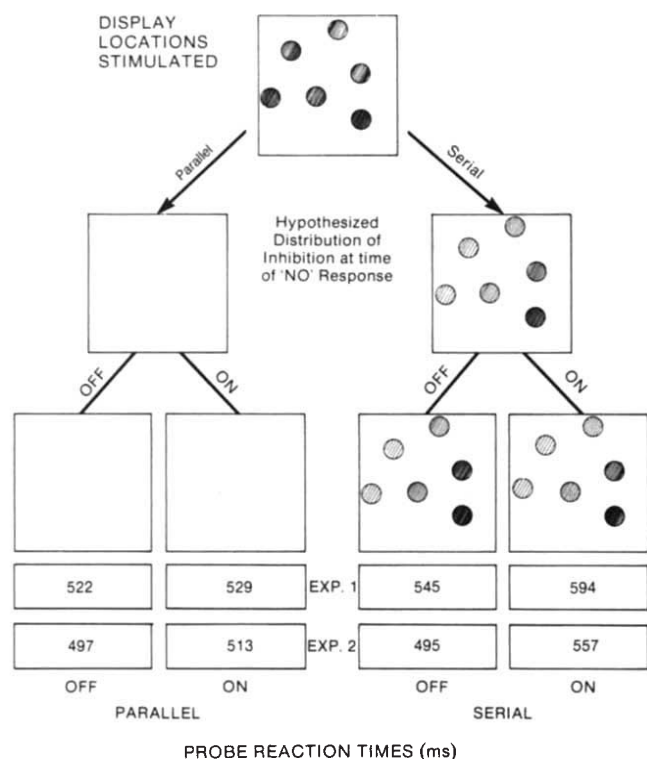
Treisman<sup>4,13,14</sup> has developed a theory of visual search in which basic features in an array of stimulus objects are detected pre-attentively (parallel search) whereas the detection of non-basic features or conjunctions of basic features requires the serial allocation of attention to each potential target (serial search). In parallel search the time to determine whether the target is present or absent is relatively unaffected by the number



**Fig. 1** Examples of search displays with set size equal to 6 are shown (these samples are not to scale). At the viewing distance used ( $\sim 50$  cm) the diameter of a circular stimulus was  $0.8^\circ$  and the maximum dimension of the rectangular grid containing the stimuli was  $5^\circ \times 6^\circ$ . Within each of the four panels, the left display is a target-present trial and the right display a corresponding target-absent trial. Search data are shown in the bottom panels for each experiment separately. Since there were no differences between the line and gap displays in either experiment the results from these have been combined. Data from the displays chosen to induce serial search are drawn in solid lines, from those chosen to induce parallel search in dotted lines. Closed symbols, data from target-present trials; open symbols, data from target-absent trials.

of displayed items. In serial search reaction time is a linearly increasing function of the number of displayed items, and the slope of the function for target-present trials is half that of the slope for target-absent trials. This is precisely the pattern that would be expected if search was serial and self-terminating. How does the serial search mechanism keep track of where attention has been, so that it does not return there again? Inhibition of return might help perform this function, perhaps in conjunction with consciously directed search strategies.

To test this idea we presented observers with serial and parallel search tasks, followed by a speeded luminance detection task. The luminance probe was either placed at a location where there had previously been a display item (on-probe) or at a previously empty location (off-probe). In serial search if the presumed allocation of attention to each item is followed by inhibition of return, then detection of on-probes should be delayed compared with off-probes. The parallel search condition, in which the target is detected pre-attentively, serves as a control or baseline for most other factors. These would include masking (from the surrounding display item following on-probes or nearby items following off-probes) and expectancy effects (due to the use of the display items as cues to possible probe locations) which should be approximately equal in the parallel and serial search conditions. If on-probes are detected less rapidly than off-probes in serial but not parallel search, then this would be strong evidence for the functional role that has been attributed to inhibition of return. Visual attention can be shifted both overtly as when we make eye movements to foveate a region, and covertly without any shift in gaze<sup>6</sup>. During serial search both



**Fig. 2** The top three sets of squares indicate the relationship between the display conditions and the proposal that inhibition of return operates during serial search. Probe reaction times following target-absent trials (collapsed across set size and gap versus line displays) from the two experiments are indicated in the bottom two lines. The critical differences (Fisher's LSD) are 43.5 ms and 40.9 ms for experiments 1 and 2, respectively.

these mechanisms are normally used. Because the inhibition effect is registered in environmental rather than retinal coordinates, it is neither essential nor desirable to prevent eye movements. In future work the impact of eye movements will be directly explored.

In two experiments observers searched for targets in displays (Fig. 1) that we expected would elicit parallel and serial search<sup>4,14</sup>. Each experimental block of 192 trials used only one of the display types (parallel lines, parallel gaps, serial lines or serial gaps), and consisted of an equal number of target-present and target-absent trials with each display size (2, 6 and 10 items) occurring equally often in each condition. Each trial began with the presentation of a plus sign, for visual fixation, in the centre of a display oscilloscope. After 1 s an array was presented and the observer pressed one of two keys with his right hand to indicate whether the display contained a target. In half these trials (representing each combination of display size and target-present/absent), the response to the target search (which terminated the search display) was followed, 60 ms later, by presentation of a luminance detection probe consisting of a single pixel illuminated on the display monitor either in the middle of a location that had been occupied by a display item (on-probe), or at a previously empty location (off-probe) (Fig. 2). The subject rapidly pressed a hand-held button with his left thumb extinguishing the luminance probe as soon as he was aware of it. In experiment 1, 14 observers were tested in a single session with one serial and one parallel block (with lines or gaps varied between subject). In experiment 2 a smaller group (five) of more experienced observers were tested in 14 blocks over four sessions (with all four display types varied between blocks).

Examination of search reaction time and accuracy (Fig. 1) reveals that for both levels of practice we obtained patterns quite similar to those reported by Treisman<sup>4,14</sup>. The serial search displays yield much steeper slopes than the parallel displays

and in the serial search conditions the slope for target-present trials is about half that of the slope for target-absent trials. The small effect of display size in the parallel search condition of experiment 1 is not atypical<sup>4,14</sup>, and is probably due to the relatively low level of practice in that experiment and a tendency for subjects to engage in serial search on a small proportion of the trials.

On any given target-present trial we cannot know which locations are attended before target detection (though on average one half will be). Our analysis of the probe latencies will therefore emphasize the target-absent trials for which, according to Treisman's model<sup>14</sup>, the location of every item is attended (Fig. 2). In both experiments the serial-on probes were responded to more slowly than the serial-off probes and either type of parallel probe (Fig. 2), with no significant difference between the latter three conditions. This pattern was statistically independent of set size. Moreover, the pattern of probe performance following serial and parallel search conditions with equivalent latencies (serial set size 2 and parallel set size 6) was nearly identical to that shown in Fig. 2 (collapsed across the two experiments: serial/on, 561; serial/off, 500; parallel/on, 523; parallel/off, 513). This demonstrates that the overall difference in exposure duration and/or search reaction time between the serial and parallel search tasks is not responsible for the observed pattern of probe reaction time. The parallel search condition can provide a baseline for assessing the effect of factors other than inhibition of return on probe reaction time. The (nonsignificant) differences between on and off probe reaction time in the parallel search condition were subtracted from the serial on/off difference to provide a purer measure of the magnitude of inhibition of return in our search task. On target-absent trials this difference score was 44.5 ms in experiment 1 and 46 ms in experiment 2. On target-present trials an effect that is half this size would be expected, an expectation that is approximately confirmed (25.5 ms in experiment 1, 28 ms in experiment 2). The pattern of results is precisely what would be predicted by the hypothesis that inhibition of return operates in serial search and provides strong support for the view that inhibition of return functions to make visual search more efficient.

Broadbent<sup>15</sup> describes a parallel activation model which does not assume a spatially allocated attentional system, and which can account for a variety of latency/search patterns (including serial exhaustive, serial self-terminating, and parallel) with minor parameter adjustments. Our study, however, provides converging evidence for the notion of spatial allocation of attention with some displays, and therefore appears inconsistent with parallel models like Broadbent's, which specifically avoid such an assumption. Finally it is interesting to consider that a similar inhibitory tagging system may operate within neuro-cognitive maps that mediate foraging behaviour in other species<sup>16,17</sup>.

Received 7 March; accepted 20 June 1988.

1. Posner, M. I. & Cohen, Y. in *Attention & Performance X: Control of Language Processes* (eds Bouma, H. & Bouwhuis, D.) 531-556 (Lawrence Erlbaum, Hillsdale, 1984).
2. Posner, M. I., Rafal, R. D., Choate, L. S. & Vaughan, J. *Cognitive Neuropsychol.* **2**, 211-228 (1985).
3. Koch, C. & Ullman, S. *Hum. Neurobiol.* **4**, 219-227 (1985).
4. Treisman, A. *Scient. Am.* **255**, 106-115 (1986).
5. Crick, F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4586-4590 (1984).
6. Posner, M. I. *Q. J. exp. Psychol.* **32**, 3-25 (1980).
7. Briand, K. & Klein, R. M. *J. exp. Psychol., hum. Percept. Perform.* **13**, 224-228 (1987).
8. Eriksen, W. & Hoffman, J. E. *Percept. & Psychophys.* **11**, 169 (1972).
9. Jonides, J. in *Attention & Performance IX* (eds Long, J. & Baddeley, A.) 187-204 (Lawrence Erlbaum, Hillsdale, 1981).
10. Sagi, D. & Julesz, B. *Nature* **321**, 693-695 (1986).
11. Maylor, E. A. & Hockey, R. J. *J. exp. Psychol., hum. Percept. Perform.* **11**, 777-787 (1985).
12. Maylor, E. A. in *Attention & Performance XI* (eds Posner, M. I. & Marin, O.) 189-207 (Lawrence Erlbaum, Hillsdale, 1985).
13. Treisman, A. & Gelade, G. *Cognitive Psychol.* **12**, 97-136 (1980).
14. Treisman, A. & Souter, J. *J. exp. Psychol., hum. Percept. Perform.* **12**, 3-17 (1986).
15. Broadbent, D. in *Modelling Cognition* (ed. Morris, P. E.) (Wiley, London, in the press).
16. O'Keefe, K. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford University Press, Oxford, 1978).
17. Olton, D. S., Handelman, G. E. & Walker, J. A. in *Foraging Behavior: Ecological, Ethological and Psychological Approaches* (eds Kamil, A. C. & Sargent, T. D.) 333 (Garland, New York, 1981).



## Creatine accelerates the circadian clock in a unicellular alga

Till Roenneberg, Hideshi Nakamura\*  
& J. Woodland Hastings

Departments of Cellular and Developmental Biology and Chemistry,\* Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

The circadian clock is considered to be a universal feature of eucaryotic organisms, controlling the occurrence and rates of many different aspects of life, ranging from single enzymatic reactions and metabolism to complex behaviours such as activity and rest<sup>1</sup>. Although the nature of the underlying cellular/biochemical oscillator is still unknown<sup>2</sup>, many substances are known to influence either phase or period of circadian rhythms in different organisms. These include  $D_2O$ <sup>3,4</sup>, electrolytes<sup>5</sup> and ion channel inhibitors<sup>5,6</sup>, small organic molecules such as alcohols<sup>7,8</sup> and aldehydes<sup>9</sup>, inhibitors of protein synthesis<sup>10-12</sup> and amino-acid analogues<sup>13,14</sup>. Certain transmitter<sup>15</sup> and neurochemical drugs<sup>16,17</sup> also influence the circadian clock in higher animals. We report here that the period of free-running circadian rhythms in the unicellular marine alga *Gonyaulax polyedra* is shortened by extracts from mammalian cells. The effect is dose-dependent, accelerating the circadian clock by as much as 4 hours per day. The substance responsible for this effect has been isolated from bovine muscle and identified as creatine. Authentic creatine has identical biological effects at micromolar concentrations and is known in animal systems for its involvement in cellular energy metabolism. A period shortening substance with similar chemical properties is also present in extracts of *Gonyaulax* itself.

In *Gonyaulax*, several functions exhibit circadian rhythmicity<sup>1</sup>, including motility, cell division, photosynthesis and bioluminescence. Under constant conditions, these persist for several weeks with a precise period<sup>18,19</sup>. The bottom trace of Fig. 1a shows a glow rhythm for nine days of constant conditions (period:  $23.3 \pm 0.3$  h).

The effect of creatine on the circadian period in *Gonyaulax*, discovered as an effect of a substance in extracts of bovine brain, is illustrated by the upper traces in Fig. 1a. When times of the glow peaks are replotted on consecutive days, the slopes of the linear regressions are proportional to period length (Fig. 1b). At the highest concentration, the period is  $\sim 18.8$  h, and the cells exhibit this maximal response for more than nine days. At lower concentrations the effects appear biphasic: the maximal effect occurs at the outset and continues for a dose-dependent time, followed by some intermediate period length, which is also dose related. A prolonged period-shortening effect does not require the continuous presence of the extract in the medium. As shown in Fig. 2, the effect of a 24-h exposure to  $1 \mu\text{l}$  per ml is comparable to a continuous exposure to  $0.4 \mu\text{l}$  per ml.

Partially purified extracts from several other cells and tissues, and also extracts from *Gonyaulax* itself, lead to the similar results: compared to the control, periods become shorter with higher concentrations, by as much as 4 h per day (Fig. 3). The period-shortening substance was further purified from bovine muscle and identified as creatine. The chromatographic behaviour and physico-chemical properties of the active compound are compared with those of authentic creatine in Fig. 4.

Creatine shortens the circadian period substantially with no evident adverse effects below concentrations of  $20 \mu\text{M}$  (for continuous exposure); below  $2 \mu\text{M}$  it has little or no activity (Fig. 3, inset). The initial suppression of luminescence at the higher concentrations in Fig. 1a is not due to general cellular impairment or inhibition: the cells' motility is actually increased by creatine during this time, and the algae remain active throughout the night phase, when they are normally quiescent.

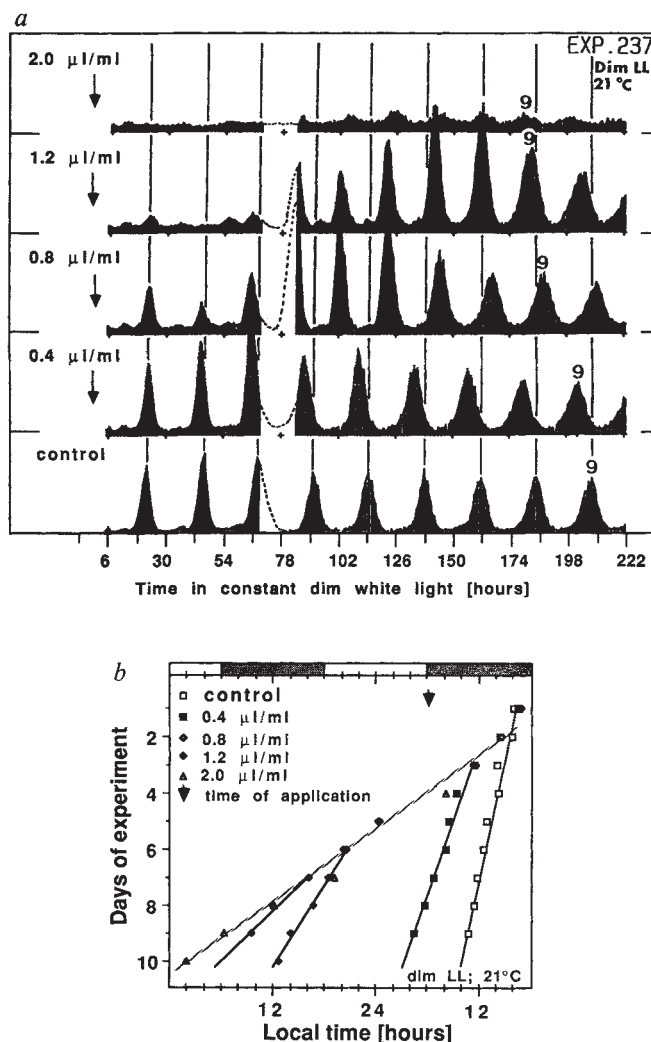
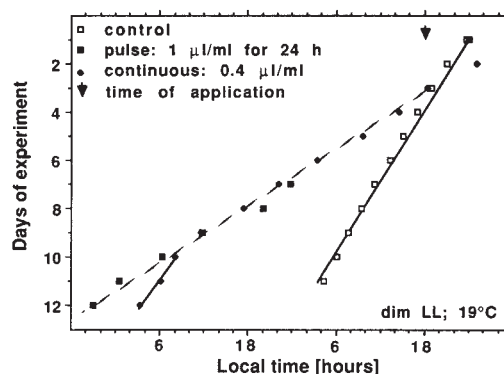


Fig. 1 a, The effect of different amounts of partially purified bovine brain extract on the glow rhythms of *Gonyaulax*, added at the time indicated (arrow). Between hours 69 and 84 the data collection failed due to computer error. Vertical lines are drawn at the times when glow maxima occurred in the control culture and the ninth glow peaks are indicated. b, Times of glow peaks on consecutive days are plotted and period lengths are calculated by linear regression (arrow = time when extracts were added). The light cycle before release into dim light is indicated at the top. **Methods.** A detailed description is given elsewhere<sup>20</sup>. Cells were grown at  $21^\circ\text{C}$  under a light-dark cycle (12L:12D,  $100 \mu\text{E} \times \text{m}^{-2} \text{s}^{-1}$ ). Aliquots of 10 ml ( $3,000 \text{ cells ml}^{-1}$ ) were dispensed in scintillation vials and transferred to constant conditions of dim light ( $40 \mu\text{E} \times \text{m}^{-2} \text{s}^{-1}$ ) and temperature ( $21^\circ\text{C}$ ). The bioluminescence of each culture was measured for 40 s every 45 min with a mobile photomultiplier.

This could be the reason for the decreased bioluminescence during the subjective night (night phase under constant conditions). The diverse effects on the circadian clock, as well as on bioluminescence and motility, are the same for creatine and extracts.

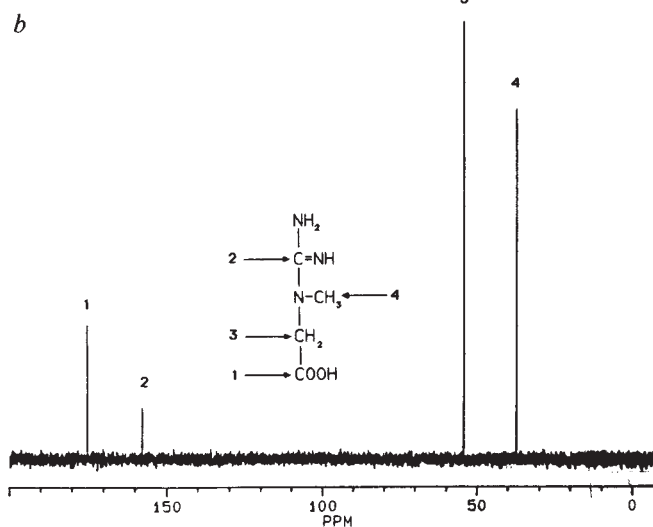
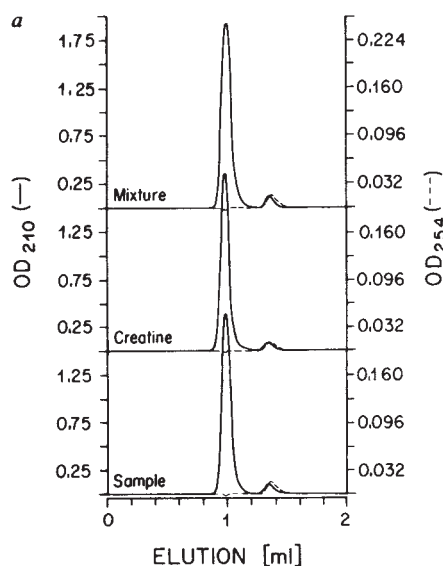
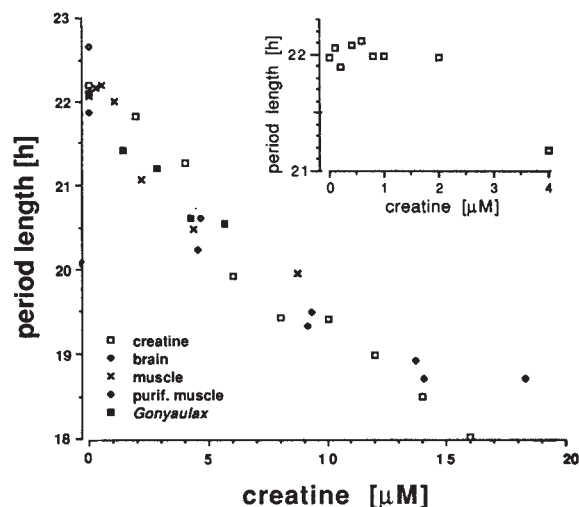
The biochemical mechanism by which creatine influences the period length of the circadian clock in this photosynthetic, motile alga is not evident. The results of uptake measurements and pulse experiments suggest an active uptake and an intracellular site of action. However, the presence of creatine in a unicellular and/or a photosynthetic organism has, to our knowledge, not been previously reported. The possibility of creatine acting as an analogue to a molecule endogenous to the algae can not be

**Fig. 2** Continuous and time-limited exposure to partially purified bovine brain extract lead to similar effects. 24 h after the addition of  $1 \mu\text{l}$  per ml extract, cells were filtered, washed and resuspended in fresh medium.  $0.4 \mu\text{l}$  per ml of the extract was added to a separate culture at the same time (arrow), but the medium was not exchanged. Glow peaks are plotted as described for Fig. 1b. In the case of the 24-h exposure, no glow peaks could be measured for several days after the addition, due to the suppression of the bioluminescence (as seen in Fig. 1a). The effect of the 24-hour exposure still prevailed after 10 days, whereas the effect of the continuous presence diminishes after 9 days.



**Fig. 3** The period-shortening effects of partially purified bovine brain and muscle extracts, homogenate of *Gonyaulax* cells as well as the purified substance and authentic creatine are compared. The shortening of the period due to the different extracts is normalized to the effect of creatine. Period lengths are calculated as an average over the time period measured (8–10 days) and thus include both components of a sometimes biphasic response (see Fig. 1b). The threshold concentration of creatine was determined more accurately in a separate experiment (see inset). Extracts of rat brain and muscle as well as homogenates of *Acanthamoeba castellanii* also lead to shorter periods, but extracts of bacteria show no effect.

**Methods.** The active compound used for physico-chemical identification (see legend to Fig. 4) was purified from bovine muscle extract: ground beef was homogenized in iso-propanol (1 mg per g) and centrifuged (Sorvall SS-39 29,000g for 20 min). The active compound was extracted from the supernatant with silica beads and eluted with double distilled  $\text{H}_2\text{O}$ . The eluate was concentrated to a saturated solution and separated further on a G-10 column (Sephadex in double distilled  $\text{H}_2\text{O}$ ). Fractions were tested for activity (see methods, Fig. 1). Active fractions were further purified with the help of HPLC ( $\text{C}_{18}$ -Ultrasphere in double distilled  $\text{H}_2\text{O}$ ).



**Figure 4** HPLC comparison and  $^{13}\text{C}$ -NMR spectrum of natural and authentic creatine. *a*, The purified sample, authentic creatine and a mixture of both were analysed by HPLC (Ultrapore  $\text{C}_3$  on Beckman 110B eluted in double distilled  $\text{H}_2\text{O}$  at  $0.2 \text{ ml min}^{-1}$ ; detection at 210 and 254 nm, Beckman models 163 and 153). Creatine absorbs only at 210 nm and coelutes with the purified substance on four different HPLC columns ( $\text{C}_{18}$ ,  $\text{C}_3$ , silica, 3000PW) under various conditions (pH and solvent). The second, smaller absorption peak in all three profiles is due to creatinine (judged by the elution volume of an authentic sample; creatinine absorbs both at 210 and 254 nm). Creatinine affects the circadian period only at concentrations  $\sim 6$  times higher than creatine, which can be explained by the existence of an equilibrium between creatine and creatinine in water. Creatine phosphate shows no biological effects when added to the culture medium, but presumably can not enter the cell. *b*, 125 MHz  $^{13}\text{C}$ -NMR spectrum of a mixture of the purified sample and authentic creatine was recorded on a Bruker AM500 instrument in  $\text{D}_2\text{O}$ . The carbon groups in the chemical structure of creatine are numbered in relationship to the signals. The physico-chemical data: Fast Atom Bombardment MS:  $m/z$  154 ( $\text{M} + \text{Na}^+$ ), 132 ( $\text{M} + \text{H}^+$ );  $^1\text{H}$ -NMR: (500 MHz in  $\text{D}_2\text{O}$ ):  $\delta$  p.p.m. 2.92 (s, 3H,  $\text{NCH}_3$ ), 3.82 (s, 2H,  $\text{NCH}_2\text{CO}$ );  $^{13}\text{C}$ -NMR (125 MHz in  $\text{D}_2\text{O}$ ):  $\delta$  p.p.m. 37.6 (q,  $\text{NCH}_3$ ), 54.6 (t,  $\text{NCH}_2\text{CO}$ ), 157.8 (s,  $\text{N}-\text{C}(=\text{NH})-\text{N}$ ), 175.3 (s,  $\text{COOH}$ ) (reference: dioxane  $\delta_c$  67.3). For both NMRs a comparison was made with authentic creatine (Katayama Chemicals, Osaka, Japan).



ruled out as long as the identity of the period-shortening substance extracted from *Gonyaulax* remains to be established. As a nitrogen-rich compound, creatine could exert its circadian activity through its guanidino-group (see structure in Fig. 4b). Yet, guanidine and arginine, as well as agmatine, show no effects when present in the culture medium.

A further possibility for the biochemical mechanism of this effect could be related to the known function of creatine in animal cells, as a storage molecule for high-energy phosphate. Recent investigations in muscle and sperm show that creatine also serves as a high-energy phosphate shuttle molecule between mitochondria and different energy-consuming sites, such as myosin and dynein<sup>21,22</sup>. Furthermore, creatine is known to stimulate ADP-regulated respiration in mitochondria<sup>23,24</sup>, leading to an increase of the cellular high-energy phosphate pool. But the existence of a creatine kinase in *Gonyaulax* is not known.

It has been proposed that the high-energy phosphate pool of the creatine system might serve as a buffer in the periodic energy consumption of heart cells<sup>24</sup>. Alternating levels of energy production and utilization, on a daily basis, are common cellular features and reflect the 'time-ecological' usage of fluctuations in the environment. In plants<sup>25</sup> and fungi<sup>26</sup> the energy charge<sup>27</sup> undergoes a pronounced daily rhythm, even under constant conditions, and energy metabolism itself may be involved in the generation of the circadian oscillation<sup>28</sup>. Moreover, creatine can be actively taken up by cells with the help of specific ports in the plasma membrane<sup>29,30</sup>.

Whether creatine as such or some metabolic product thereof acts on the circadian clock in *Gonyaulax* or whether creatine is analogous to an endogenous molecule remains to be investigated. The fact that *Gonyaulax* cells continue to be highly active throughout the circadian cycle when creatine is added to the culture medium could indicate an effect on energy metabolism. This could result in a higher circadian frequency if creatine can 'buffer' the daily fluctuations in photosynthetic energy production<sup>1,19</sup> and thus increase the overall level of available cellular energy.

This work was supported by grants of the NIH, the NSF and the Deutsche Forschungsgemeinschaft. We thank Dr Neil R. Krieger for helpful discussions and Dr Y. Kishi for his help in identifying the structure of the substance.

Received 6 May; accepted 10 June 1988.

- Johnson, C. H. & Hastings, J. W. *Am. Scient.* **74**, 29–36 (1986).
- The Molecular Basis of Circadian Rhythms* (eds Hastings, J. W. & Schweiger, H.-G.) (Dahlem Conferences, Berlin, 1976).
- Bruce, V. G. & Pittendrigh, C. S. *J. Cell. comp. Physiol.* **56**, 25–31 (1960).
- McDaniel, M., Sulzman, F. M. & Hastings, J. W. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4389–4391 (1974).
- Schmid, H. P. & Engelmann, W. *J. interdiscipl. Cycle Res.* **16**, 324 (1986).
- Kondo, T. *J. biol. Rhyth.* (submitted).
- Enright, J. T. *J. comp. Physiol.* **75**, 332–346 (1971).
- Taylor, W., Gooch, V. D. & Hastings, J. W. *J. comp. Physiol. B* **130**, 355–358 (1979).
- Taylor, W. & Hastings, J. W. *J. comp. Physiol. B* **130**, 359–362 (1979).
- Karakashian, M. W. & Hastings, J. W. *J. gen. Physiol.* **47**, 1–12 (1963).
- Feldman, J. F. *Proc. natn. Acad. Sci. U.S.A.* **57**, 1080–1087 (1967).
- Rothman, B. & Strumwasser, F. *J. gen. Physiol.* **68**, 359–384 (1976).
- Hastings, J. W., Broda, H. & Johnson, C. H. in *Temporal Order* (eds Rensing, L. & Jäger, N. I.) 213–221 (Springer, Berlin, 1985).
- Kondo, T. *Pl. Physiol.* (submitted).
- Corrent, G., & Eskin, A. *Am. J. Physiol.* **11**, R333–338 (1982).
- Turek, F. W. & Losee-Olsen, S. *Nature* **321**, 167–168 (1986).
- Ralph, M. R. & Menaker, M. *Brain Res.* **325**, 362–365 (1985).
- Njus, D., Gooch, V. & Hastings, J. W. *Cell Biophys.* **3**, 223–231 (1981).
- McMurry, L. & Hastings, J. W. *Science* **175**, 1137–1139 (1972).
- Taylor, W., Wilson, S., Presswood, R. P. & Hastings, J. W. *J. interdiscipl. Cycle Res.* **13**, 71–79 (1982).
- Tombes, R. M. & Shapiro, B. M. *Cell* **41**, 325–334 (1985).
- Bessman, S. P. *Analyt. Biochem.* **161**, 519–523 (1987).
- Bessman, S. P. & Fonjo, A. *Biochem. biophys. Res. Commun.* **22**, 597–602 (1966).
- Jacobus, W. E. & Lehninger, A. L. *J. biol. Chem.* **248**, 4803 (1973).
- Wagner, E., Tetzner, J., Haertle, U. & Deitzer, G. F. *Ber. Deutsch. Bot. Ges.* **87**, 291–302 (1974).
- Delmer, D. P. & Brody, S. *J. Bact.* **121**, 548–553 (1975).
- Atkinson, D. E. *Adv. Enzyme Regul.* **9**, 207–219 (1971).
- Wagner, E. in *The Molecular Basis of Circadian Rhythms* (eds Hastings, J. W. & Schweiger, H.-G.) 215–238 (Dahlem Conferences, Berlin, 1976).
- Loike, J. D., Somes, M. & Silverstein, S. C. *Am. J. Physiol.* **251**, C 128–135 (1986).
- Loike, J. D., Zalutsky, D. L., Kanack, E., Miranda, A. F. & Silverstein, S. C. *Proc. natn. Acad. Sci. U.S.A.* (in the press).

## Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes

Ernest G. Peralta\*, Avi Ashkenazi\*, John W. Winslow†, J. Ramachandran‡§ & Daniel J. Capon\*

Departments of Molecular Biology\* and Developmental Biology†, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, USA

‡ Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

Muscarinic acetylcholine receptors (mAChRs), like many other neurotransmitter and hormone receptors, transduce agonist signals by activating G proteins to regulate ion channel activity and the generation of second messengers via the phosphoinositide (PI) and adenylyl cyclase systems<sup>1,2</sup>. Human mAChRs are a family of at least four gene products which have distinct primary structures, ligand-binding properties and patterns of tissue-specific expression<sup>3</sup>. To examine the question of whether functional differences exist between multiple receptor subtypes, we have investigated the ability of each subtype to regulate PI hydrolysis and adenylyl cyclase when expressed individually in a cell lacking endogenous mAChRs. We show that the HM2 and HM3 mAChRs efficiently inhibit adenylyl cyclase activity but poorly activate PI hydrolysis. In contrast, the HM1 and HM4 mAChRs strongly activate PI hydrolysis, but do not inhibit adenylyl cyclase, and in fact can substantially elevate cAMP levels. Interestingly, the subtypes that we find to be functionally similar are also more similar in sequence. Our results indicate that the different receptor subtypes are functionally specialized.

We previously described the transient functional expression of recombinant human mAChRs after transfection of human embryonic kidney cells<sup>3</sup>. To facilitate effector coupling studies, we established cell lines stably transfected with expression vectors for each of four human mAChR subtypes by cotransfection with a plasmid conferring neomycin resistance. Following selection of stable transformants with the antibiotic G418, cell lines expressing specific receptor numbers, determined by whole cell binding to the muscarinic antagonist [<sup>3</sup>H]quinuclidinyl benzilate (QNB), were isolated by single-cell cloning. As summarized in Table 1, cell lines were obtained which expressed comparable levels of the four human mAChR subtypes (121,000–199,000 receptors per cell); by contrast, untransfected cells expressed fewer than 400 QNB binding sites per cell.

Previous studies using selective muscarinic ligands to discriminate between different mAChR subtypes have revealed correlations between the occupancy of a particular subtype by agonist and the preferential activation of a specific cellular effector. For example, the muscarinic antagonist pirenzepine, which distinguishes between high-affinity M1 and low-affinity M2 receptors<sup>4,5</sup>, displays a differential potency in inhibiting certain muscarinic agonist-induced biochemical responses, suggesting the preferential coupling of rat brain M1 receptors and rat cardiac M2 receptors to the activation of PI hydrolysis and the inhibition of adenylyl cyclase, respectively<sup>2,6,7</sup>. We thus began our study of the functional differences between mAChR subtypes by comparing the ability of the human M1 (HM1) and M2 (HM2) subtypes<sup>3</sup> to activate PI hydrolysis when expressed in the same cell type. Dose response analysis of the agonist-induced PI hydrolysis mediated by cell lines expressing comparable levels of the HM1 and HM2 subtypes revealed that half maximal stimulation (EC<sub>50</sub>) occurred at 1 × 10<sup>-6</sup> and 1.6 × 10<sup>-5</sup> M concentrations of the acetylcholine analogue carbachol, respectively, and that a maximal response was evoked by each

§ Present address: Neurex Corporation, 3760 Haven Avenue, Menlo Park, CA 94025, USA.

**Table 1** The effect of carbachol on inositol phosphate and cAMP formation

| mAChR cell lines | <sup>[3H]</sup> QNB sites per cell (×10 <sup>-3</sup> ) | <sup>[3H]</sup> QNB K <sub>D</sub> (pM) | Maximal accumulation of inositol phosphates (c.p.m. per 10 <sup>6</sup> cells) |            |                 |            |                 | EC <sub>50</sub> Carbachol (M) |                          |                       |
|------------------|---------------------------------------------------------|-----------------------------------------|--------------------------------------------------------------------------------|------------|-----------------|------------|-----------------|--------------------------------|--------------------------|-----------------------|
|                  |                                                         |                                         | IP <sub>1</sub>                                                                | Time (min) | IP <sub>2</sub> | Time (min) | IP <sub>3</sub> | Time (min)                     | IP <sub>1</sub> (30 min) | cAMP inhibition       |
| Control          | 0.3±0.1                                                 | 19±8                                    | 240±11                                                                         | 1          | 79±4            | 5          | 80±10           | 0.25, 1                        | 7.0×10 <sup>-6</sup>     | 14.0×10 <sup>-6</sup> |
| HM1              | 141±7.8                                                 | 58±12                                   | 8,906±105                                                                      | 30         | 2,984±25        | 20         | 1,378±26        | 20                             | 1.0×10 <sup>-6</sup>     | 2.0×10 <sup>-6</sup>  |
| HM2              | 121±3.3                                                 | 44±1                                    | 1,528±343                                                                      | 30         | 154±19          | 2          | 247±13          | 0.5-2                          | 16.0×10 <sup>-6</sup>    | 3.7×10 <sup>-8</sup>  |
| HM3              | 24±0.8                                                  | 17±5                                    | 359±5                                                                          | 20         | 80±12           | 1-5        | 207±50          | 1                              |                          |                       |
| HM3              | 168±28.6                                                | 75±24                                   | 477±85                                                                         | 30         | 158±17          | 0.5        | 195±26          | 0.5                            | 4.0×10 <sup>-6</sup>     | 1.3×10 <sup>-8</sup>  |
| HM3              | 306±30.7                                                | 94±15                                   | 479±2                                                                          | 30         | 177±50          | 1          | 219±46          | 0.5, 1                         |                          |                       |
| HM4              | 13±0.3                                                  | 15±1                                    | 1,075±145                                                                      | 30         | 293±30          | 20         | 239±58          | 1                              |                          |                       |
| HM4              | 199±3.4                                                 | 19±3                                    | 14,112±1,621                                                                   | 30         | 4,075±1,093     | 5          | 1,390±259       | 5                              | 1.6×10 <sup>-6</sup>     | 8.0×10 <sup>-6</sup>  |

Analysis of  $[^3\text{H}]\text{QNB}$  binding and the effect of carbachol on inositol phosphate and cAMP formation in transfected cell lines expressing individual human mAChR subtypes. Scatchard analysis of whole-cell saturation binding experiments with the muscarinic antagonist  $[^3\text{H}]\text{QNB}$  was used to determine mAChR expression levels ( $[^3\text{H}]\text{QNB}$  binding sites per cell) and apparent dissociation constants ( $K_D$ ) as described previously<sup>3</sup>; the data represent the means  $\pm$  s.d. of two determinations performed in duplicate. The maximal accumulation of inositol phosphates was determined at the incubation time points resulting in the highest level of each hydrolysis product. The level of each inositol phosphate generated in the absence of carbachol (basal level) was subtracted and the data represent the means of two experiments  $\pm$  s.d. for the HM1, HM3 (306,000 receptors), and control cell lines or the mean  $\pm$  s.e.m. of three experiments performed in duplicate. The experimental error in measuring basal levels was incorporated into these data. Differences between the basal levels varied by less than 30% for all cell lines. For some cell lines, the maximal accumulation of  $\text{IP}_2$  or  $\text{IP}_3$  occurred over a small range of incubation times between replicate experiments; in those cases the data represent the means of the maximal values obtained from each experiment within the indicated time interval. The  $\text{EC}_{50}$  for agonist-induced PI hydrolysis was determined by dose response analysis with carbachol by measuring  $\text{IP}_1$  formation 30 min after addition of agonist; responses with  $10^{-3}$  M doses of carbachol were determined in the presence or absence of  $10 \mu\text{M}$  atropine (not shown). The  $\text{EC}_{50}$  for carbachol-induced inhibition or stimulation of cAMP levels was determined from dose response analysis (see Fig. 3). Human kidney cells were cotransfected with a vector directing the expression of each human mAChR subtype and the plasmid pRSVNeo as described<sup>3,28,29</sup>. Clonal cell lines were selected in the presence of 0.5 mg per ml G418 and screened for mAChR expression. Control cells were cotransfected with pRSVNeo and an irrelevant plasmid and selected with 0.5 mg per ml G418. The homogeneous, high-affinity  $[^3\text{H}]\text{QNB}$  binding displayed by each cell line was found to be similar to that reported for mAChRs expressed in natural tissues<sup>3</sup>.

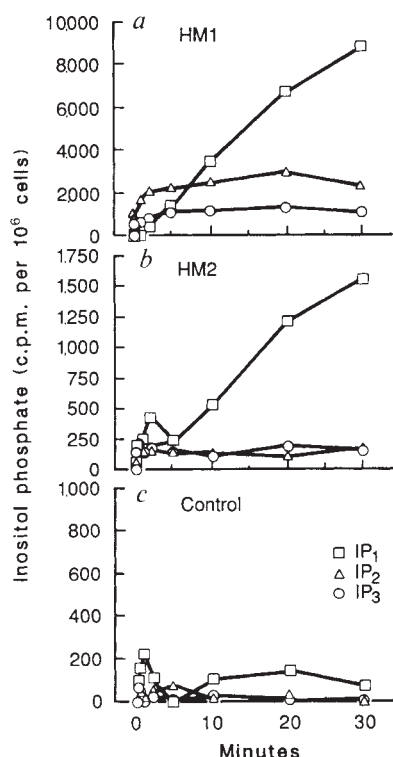
subtype at  $10^{-4}$  to  $10^{-3}$  M carbachol; furthermore, each response was completely inhibited by the non-selective muscarinic antagonist atropine (Table 1). Following treatment of cell lines expressing similar numbers of HM1 and HM2 receptors with a saturating dose of  $10^{-3}$  M carbachol, kinetic analysis of the accumulation of inositol mono-, bis-, and triphosphates ( $\text{IP}_1$ ,  $\text{IP}_2$ ,  $\text{IP}_3$ ) revealed significant differences in both the duration and magnitude of the PI response mediated by these two mAChR subtypes (Fig. 1). Carbachol treatment of HM1-expressing cells (141,000 receptors per cell) resulted in a substantial stimulation in the formation of all three inositol phosphates;  $\text{IP}_2$  and  $\text{IP}_3$  levels increased rapidly within the first minute and were maintained for 20 min following agonist exposure, whereas  $\text{IP}_1$  levels increased linearly for at least 30 min (Fig. 1). In contrast, carbachol treatment of HM2-expressing cells (121,000 receptors per cell) led to a rise in  $\text{IP}_1$  accumulation that was much smaller, though significant compared to control. The levels of  $\text{IP}_2$  and  $\text{IP}_3$  increased only slightly in HM2-expressing cells relative to untransfected cells (Fig. 1). Unlike the prolonged increase in  $\text{IP}_2$  and  $\text{IP}_3$  formation stimulated by carbachol in HM1-expressing cells,  $\text{IP}_2$  and  $\text{IP}_3$  levels reached a maximum within the first two minutes following addition of the agonist to HM2-expressing or untransfected cells (Table 1).

We next examined the ability of the two novel human mAChR subtypes identified by molecular cloning to mediate PI hydrolysis. The HM3 and HM4 subtypes exhibit an affinity for pirenzepine intermediate between those of HM1 and HM2, as well as distinct affinities for a range of other muscarinic agonists and antagonists<sup>3</sup>. Like HM1 and HM2, both these subtypes are expressed in brain, whereas in NG108-15 neuroblastoma/glioma cells only mRNA for HM3, and in pancreas only mRNA for HM4, is found<sup>3</sup>. However, the effector coupling properties of HM3 and HM4 are unknown. We investigated responses in cells expressing levels of HM3 and HM4 similar to those in the HM1- and HM2-expressing lines, and also examined the effect of receptor number on these responses. Initial dose response analysis with cell lines expressing comparable levels of HM3 and HM4 receptors revealed that the  $\text{EC}_{50}$  for the PI response was  $4 \times 10^{-6}$  and  $1.6 \times 10^{-6}$  M carbachol, respectively, and that each response was completely inhibited by atropine (Table 1). Treatment of HM4-expressing cells (199,000 receptors per cell) with a saturating dose of  $10^{-3}$  M carbachol elicited a large stimulation

of PI hydrolysis, with each of the three inositol phosphates accumulating at a rate and final magnitude very similar to that observed for cells expressing comparable levels of HM1 receptors (Fig. 2). In cells expressing substantially fewer HM4 receptors (13,000 receptors per cell), the kinetic profile of accumulation of each inositol phosphate was comparable to that observed for cells expressing 199,000 HM4 receptors per cell; however, the magnitude of the response was proportionately smaller with respect to the receptor number (Table 1). The close correlation between HM4 receptor number and the magnitude of the PI response suggests that over this 15-fold range of expression, the HM4 receptor number is the limiting factor in the pathway generating the response. In contrast to the large increase in PI hydrolysis mediated by HM1 and HM4, carbachol treatment of cells expressing 24,000 receptors, 168,000 receptors or 306,000 HM3 receptors per cell resulted in only a slight stimulation in the formation of inositol phosphates over that observed with untransfected cells. Levels of  $\text{IP}_2$  and  $\text{IP}_3$  reached a maximum within two minutes following addition of agonist and gradually decayed whereas accumulation of  $\text{IP}_1$  increased at a rate smaller than that observed for the other mAChR subtypes (Fig. 2). The magnitude of the PI response mediated by HM3 was only slightly greater in cells expressing substantially higher receptor numbers; moreover, this difference was not proportional to the receptor level (Table 1). Thus, the HM3 subtype is functionally similar to HM2 in its limited ability to stimulate PI hydrolysis in these cells.

To determine whether the four human mAChR subtypes also differ in their abilities to inhibit adenylyl cyclase activity at similar receptor numbers, the effects of carbachol on intracellular levels of adenosine 3',5'-monophosphate (cAMP) were determined following direct stimulation of adenylyl cyclase by forskolin; the phosphodiesterase inhibitor isobutylmethylxanthine (IBMX) was included to ensure that changes in cAMP levels were not due to changes in phosphodiesterase activity (Fig. 3). In contrast to the limited activation of PI hydrolysis observed within cell lines expressing the HM2 or HM3 receptors, carbachol treatment of either of these two subtypes resulted in a significant inhibition in the forskolin-stimulated accumulation of intracellular cAMP (Fig. 3). Up to 66% and 74% of the stimulated levels of cAMP were inhibited by  $10^{-4}$  M carbachol in cells expressing the HM2 or HM3 subtype, respectively,



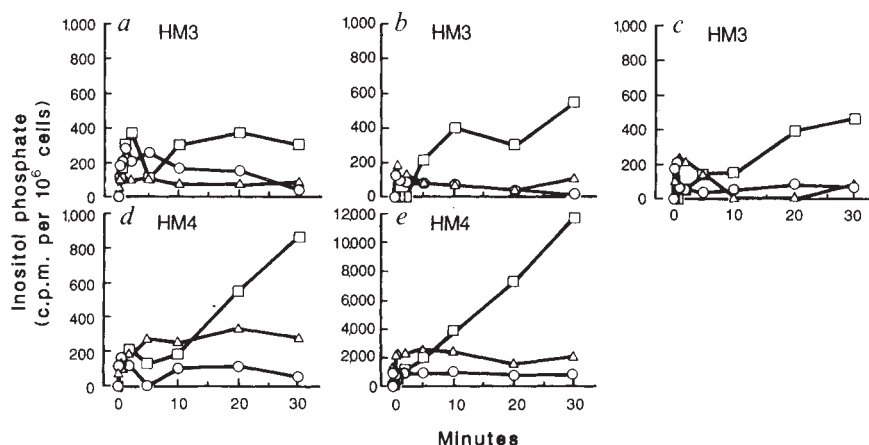


**Fig. 1** Time course of carbachol-stimulated accumulation of inositol phosphates in transfected cell lines expressing the HM1 and HM2 mAChR subtypes. *a*, Cell line expressing 141,000 HM1 receptors per cell; *b*, cell line expressing 121,000 HM2 receptors per cell; *c*, control cells expressing 300 endogenous mAChRs per cell.

**Methods.** Confluent cells were labelled with [ $^3$ H]myo-inositol (48 h,  $2 \mu\text{Ci ml}^{-1}$ ), stimulated with a saturating concentration of 1 mM carbachol for time periods of 0.25, 0.5, 1, 2, 5, 10, 20 and 30 min in the presence of 10 mM LiCl (to inhibit dephosphorylation of  $\text{IP}_1$ ), and levels of  $\text{IP}_1$  ( $\square$ ),  $\text{IP}_2$  ( $\triangle$ ) and  $\text{IP}_3$  ( $\circ$ ) were determined by anion exchange chromatography as described<sup>26,27</sup>. The basal level of each inositol phosphate was subtracted from each time point. Each cell line was tested two to four times in duplicate; results from a representative experiment are shown. The presence of the neomycin analogue G418 was not found to have an inhibitory effect on PI hydrolysis in these cell lines; experiments performed in the absence of G418 yielded similar results (not shown).

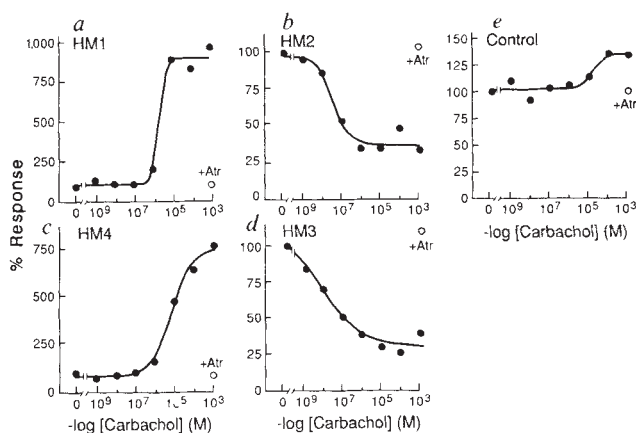
whereas the  $\text{EC}_{50}$  for the inhibition of adenylyl cyclase occurred at  $3.7 \times 10^{-8}$  M and  $1.3 \times 10^{-8}$  M carbachol, respectively (Fig. 3, Table 1). No significant inhibition of adenylyl cyclase activity above control cells was observed in cell lines expressing either HM1 or HM4 subtypes; rather, higher doses of carbachol lead to 9.8- and 7.6-fold increases in cAMP accumulation above forskolin-stimulated levels, respectively, within cell lines expressing these subtypes (Fig. 3). The  $\text{EC}_{50}$  for the stimulation of cAMP accumulation mediated by the HM1 and HM4 subtypes was  $2 \times 10^{-6}$  and  $8 \times 10^{-6}$  M carbachol, respectively (Fig. 3). The inhibition or stimulation of cAMP levels mediated by each of the four human mAChR subtypes was inhibited by the antagonist atropine (Fig. 3).

These results directly establish that the four human mAChR subtypes differentially regulate PI hydrolysis and adenylyl cyclase activity when present in the same cellular context. Although carbachol stimulation of each subtype led to increases in PI hydrolysis, stimulation of HM1 and HM4 resulted in substantially greater and more prolonged elevations of inositol phosphate levels. In addition, the 4–16-fold lower  $\text{EC}_{50}$  observed for the activation of PI hydrolysis displayed by HM1 and HM4 suggests that these two subtypes may be more efficiently coupled to PI hydrolysis than HM2 and HM3. In contrast, at levels of carbachol below those required to stimulate PI hydrolysis measurably ( $10^{-9}$  to  $10^{-7}$  M), HM2 and HM3 inhibited forskolin-stimulated cAMP levels substantially, whereas HM1 and HM4 had no significant effect. Thus, the HM1 and HM4 subtypes do not mediate the inhibition of adenylyl cyclase in these cells at agonist concentrations sufficient to evoke a substantial response by HM2 and HM3. At high levels of carbachol ( $10^{-6}$  to  $10^{-3}$  M) which elicit maximal inhibition of adenylyl cyclase by HM2 and HM3, however, HM1 and HM4 elevated cAMP beyond the forskolin-stimulated levels. Previous studies have demonstrated that cellular factors which become activated following PI hydrolysis as a result of the formation of diacylglycerol and  $\text{IP}_3$ -mediated calcium mobilization<sup>8</sup>, such as protein kinase C (PKC)<sup>9</sup> and calcium-dependent calmodulin, can stimulate adenylyl cyclase activity<sup>10,11</sup>. These observations, together with the similarities between the  $\text{EC}_{50}$  values required for the activation of PI hydrolysis and stimulation of cAMP levels by HM1 and HM4 ( $1.0 \times 10^{-6}$  and  $1.6 \times 10^{-6}$  M carbachol versus  $2 \times 10^{-6}$  and  $8 \times 10^{-6}$  M carbachol, respectively) suggest that the increases in cAMP levels mediated by HM1 and HM4 may be a secondary effect of the efficient activation of PI hydrolysis by these subtypes, but could alternatively result from the direct activation of adenylyl cyclase through a stimulatory G protein such as  $\text{G}_s$  (ref. 1). Whether the elevation of cAMP levels by



**Fig. 2** Time course of carbachol-stimulated inositol phosphate accumulation in transfected cell lines expressing the HM3 and HM4 mAChR subtypes. *a*, Cell line expressing 24,000 HM3 receptors per cell; *b*, 168,000 HM3 receptors per cell; *c*, 306,000 HM3 receptors per cell; *d*, 13,000 HM4 receptors per cell; *e*, 199,000 HM4 receptors per cell.

**Methods.** Levels of  $\text{IP}_1$  ( $\square$ ),  $\text{IP}_2$  ( $\triangle$ ) and  $\text{IP}_3$  ( $\circ$ ) were determined as described in Fig. 1. Each cell line was tested two to four times in duplicate and results from a representative experiment are shown.



**Fig. 3** Effect of carbachol on cAMP formation in transfected cell lines expressing individual human mAChR subtypes. *a*, HM1, 141,000 receptors per cell; *b*, HM2, 121,000 receptors per cell; *c*, HM4, 199,000 receptors per cell; *d*, HM3, 168,000 receptors per cell; *e*, control cells, 300 endogenous mAChRs per cell. Open circles indicate the response in the presence of 20  $\mu$ M atropine. **Methods.** Cells were equilibrated with 200  $\mu$ M IBMX and treated with 10  $\mu$ M forskolin and varying doses of carbachol for an incubation period of 30 s; cAMP levels were quantified by radioimmunoassay as described previously<sup>26</sup>. The level of cAMP in IBMX-treated cells was  $2.4 \pm 0.1$ ,  $4.1 \pm 0.1$ ,  $2.4 \pm 0.1$ , and  $2.7 \pm 0.1$  pmol per  $10^6$  cells for cell lines expressing the HM1, HM2, HM3, HM4 subtypes and control cells, respectively. Forskolin increased levels of cAMP to  $129 \pm 9$ ,  $113 \pm 4$ ,  $270 \pm 14$ ,  $57 \pm 5$  and  $118 \pm 5$  pmol per  $10^6$  cells in cell lines expressing the HM1, HM2, HM3, HM4 subtypes and control cells, respectively, in the absence of carbachol. The maximal inhibition or stimulation of cAMP levels resulted in  $1269 \pm 109$ ,  $38 \pm 3$ ,  $69 \pm 7$ ,  $429 \pm 12$  and  $135 \pm 4$  pmol per  $10^6$  cells for the HM1, HM2, HM3, HM4 and control cells, respectively.  $10^{-3}$  M doses of carbachol were tested in the presence or absence of 20  $\mu$ M atropine. The data represent the means  $\pm$  s.e.m. of four determinations at each dose of carbachol.

HM1 and HM4 observed here is of physiological significance remains to be determined; note, however, that carbachol treatment of 1321N1 human astrocytoma cells in the presence of IBMX leads to a small but significant increase in isoproterenol-stimulated cAMP levels<sup>12</sup>. Recent studies suggest that 1321N1 cells primarily express the HM4 subtype (J.W. *et al.*, unpublished data).

The efficient stimulation of PI breakdown observed here for HM1 and HM4, and their high level of mRNA expression in brain, suggest that these two subtypes may play important roles in central nervous system function through activation of PKC and other calcium-dependent kinases, as well as the potential regulation of calcium-modulated neurotransmitter release<sup>13</sup>. The subtype-specific expression of HM4 in pancreas<sup>3</sup> also suggests that this subtype may evoke the calcium-dependent enzyme secretion activities of this tissue<sup>14</sup>. The fact that the stimulation of PI hydrolysis mediated by the HM2 and HM3 subtypes is so small indicates that, if it has a physiological function, this function may be distinct from that of the large increases mediated by HM1 and HM4. The efficient inhibition of adenylyl cyclase activity by HM2 and HM3 on the other hand, indicates that these subtypes may regulate biological responses largely through this signalling pathway, in addition to the regulation of heart K channels known to be mediated by M2 mAChRs (ref. 2). In agreement with these observations, the expression of M2 and M3 mAChR mRNA in heart tissue and NG108-15 neuroblastoma  $\times$  glioma cells, respectively, is correlated with the ability of muscarinic agonists to inhibit adenylyl cyclase activity in these cells<sup>3,7,12</sup>.

The structure of the four human mAChRs like other G protein-linked receptors and visual rhodopsins is predicted to contain seven membrane-spanning regions which together may form the ligand-binding domain of this class of receptors<sup>15-19</sup>. Consistent with this model, the four human mAChR subtypes share a high degree of amino acid identity within their proposed transmembrane domains, but diverge substantially in regions predicted to reside on the cytoplasmic side of the lipid bilayer<sup>3,20</sup>. Interestingly, the greater similarity observed between the HM1 and HM4 or the HM2 and HM3 subtypes within these cytoplasmic regions correlates with the ability of these subtypes to interact preferentially with a particular second-messenger system as demonstrated here. In particular, each of these mAChR subtype pairs displays 21% amino acid identity in the large cytoplasmic loop connecting the fifth and sixth transmembrane segments, a region that is otherwise highly divergent between different mAChR subtypes<sup>3,20</sup>. Sequences conserved between the large cytoplasmic loops of HM1 and HM4 or HM2 and HM3 may thus be important in conferring the ability of these subtypes to interact more efficiently with G proteins which mediate PI hydrolysis (Gp)<sup>21</sup> or the inhibition of adenylyl cyclase activity (Gi)<sup>22</sup> within these cells.

The differential ability of human mAChR subtypes to activate PI hydrolysis and inhibit adenylyl cyclase and the previous demonstration that recombinant porcine M1 and M2 mAChRs expressed in *Xenopus* oocytes have distinct electrophysiological effects<sup>23</sup> indicate that each subtype has evolved the structural capacity to preferentially activate specific cellular responses. A similar degree of structural diversity occurs within other components of cellular signal transduction pathways, including the family of G and related proteins<sup>1</sup>, as well as the multiple forms of cAMP-dependent protein kinases and PKC<sup>24,25</sup>, suggesting that each member of these multigene families has evolved specific functions. The expression of individual mAChR subtypes in transfected cells should be a useful strategy for obtaining further insight into the ability of each subtype to evoke other cellular biochemical and electrophysiological responses, and to interact differentially with specific members of signal transduction pathways.

We thank Scot Marsters for cell culture assistance and K. Andow for preparation of the figures. This work was supported by Genentech, Inc. and by the NIH (grant to J.R.).

Received 7 March; accepted 14 June 1988.

1. Stryer, L. & Bourne, H. A. *Rev. Cell Biol.* **2**, 391-419 (1986).
2. Nathanson, N. A. *Rev. Neurosci.* **10**, 195-236 (1987).
3. Peralta, E. *et al. EMBO J.* **6**, 3923-3929 (1987).
4. Hammer, R. *et al. Nature* **283**, 90-92 (1980).
5. Birdsall, N. & Hulme, E. *Trends Pharmacol. Sci.* **4**, 459-463 (1983).
6. Watson, M. *et al. Trends Pharmacol. Sci.* **7**, 46-55 (1986).
7. Gil, D. & Wolfe, B. J. *Pharmac. exp. Ther.* **232**, 608-616 (1985).
8. Berridge, M. J. A. *Rev. Biochem.* **56**, 159-193 (1987).
9. Nishizuka, Y. *Science* **233**, 305-312 (1986).
10. Yoshimasa, T. *et al. Nature* **327**, 67-70 (1987).
11. Salter, R., Krinks, M., Klee, C. & Neer, E. J. *biol. Chem.* **256**, 9830-9833 (1981).
12. Hughes, A. & Harden, T. J. *Pharmac. exp. Ther.* **237**, 173-178 (1987).
13. Rabe, C., Delorme, E., Weight, F. J. *Pharmac. exp. Ther.* **243**, 534-541 (1987).
14. Korc, M., Ackerman, M., Roeske, W. J. *Pharmac. exp. Ther.* **232**, 608-616 (1985).
15. Peralta, E. *et al. Science* **236**, 600-605 (1987).
16. Kubo, T. *et al. Nature* **323**, 411-416 (1986).
17. Kubo, T. *et al. FEBS Lett.* **209**, 367-372 (1986).
18. Dixon, R. *et al. Nature* **321**, 75-79 (1986).
19. Nathans, J., Thomas, D., Hogness, D. *Science* **232**, 193-200 (1986).
20. Peralta, E. *et al. Trends in Pharmacol. Sci. February Suppl.* 6-11 (1988).
21. Cockcroft, S. & Gomperts, B. D. *Nature* **314**, 534-536 (1985).
22. Bray, P. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 5115-5119 (1987).
23. Fukuda, K. *et al. Nature* **327**, 623-625 (1987).
24. Uhler, M. & McKnight, G. S. *J. biol. Chem.* **262**, 15202-15207 (1987).
25. Coussens, L. *et al. Science* **233**, 859-866 (1986).
26. Ashkenazi, A. *et al. Science* **238**, 672-675 (1987).
27. Berridge, M. J., Downes, C. & Hanley, M. R. *Biochem. J.* **206**, 587-595 (1982).
28. Eaton, D. L. *et al. Biochemistry* **26**, 8343-8347 (1986).
29. Gorman, C., Padmanabam, R. & Howard, B. *Science* **221**, 551-553 (1983).
30. Schimerlik, M. & Searles, R. *Biochemistry* **19**, 3407-3413 (1980).



## Increased pH and tumorigenicity of fibroblasts expressing a yeast proton pump

Rosario Perona\* & Ramon Serrano†

\* Instituto de Investigaciones Biomedicas del CSIC, Facultad de Medicina de la Universidad Autonoma, Arzobispo Morcillo 4, 28029 Madrid, Spain

† European Molecular Biology Laboratory, Postfach 10.2209, 6900 Heidelberg, FRG

A common early response of eukaryotic cells to stimuli which activate their proliferation is an increase in intracellular pH (ref. 1). In animal cells this is caused by the activation of an  $\text{Na}^+/\text{H}^+$  exchange system<sup>2-5</sup>; in fungi and plants an  $\text{H}^+$ -pumping ATPase<sup>6</sup> is involved. The critical question is whether this intracellular alkalinization is merely coincident with the activation of cell proliferation or whether it is a regulatory signal<sup>2</sup>. To increase intracellular pH bypassing the usual physiological stimuli (growth factors, hormones etc.) alkaline media or ammonia have been used in the past<sup>2</sup>. Both approaches suffer from long-term toxicity effects and cannot be used in tumorigenic assays with whole organisms. We introduce here a more specific approach which involves expressing the gene for the yeast plasma membrane  $\text{H}^+$ -ATPase<sup>7</sup> in fibroblasts. The resulting cells have an elevated intracellular pH and acquire tumorigenic properties, suggesting that the yeast ATPase gene behaves as an oncogene in mammalian cells. These experiments support a crucial role of intracellular pH in the growth control of animal cells.

The gene for the yeast plasma membrane  $\text{H}^+$ -ATPase has been inserted into the mammalian expression plasmid pSV<sub>2</sub> under the control of the SV40 promoter<sup>8</sup>. The ATPase fragment used extended from an artificially introduced *Hind*III site 70 base pairs (bp) upstream of the initiation codon to the *Hind*III site downstream of the gene<sup>7</sup>. The resulting plasmid (pSV<sub>h</sub>ATP<sub>5</sub>) was cotransfected with plasmid pSV<sub>2</sub>neo<sup>8</sup> into mouse (NIH3T3) and monkey (Vero) fibroblasts. Three out of six neomycin-resistant clones from every cell line expressed the yeast ATPase as detected by immunofluorescence of fixed and permeabilized cells<sup>8</sup> (Fig. 1). As intact cells expressing the ATPase showed weak but reproducible labelling (data not shown), at least part of the ATPase seems to be located at the plasma membrane. All the mouse and monkey clones expressing the yeast ATPase exhibited transformed properties: they grew to high population density, with many rounded and highly refractile cells (Fig. 2), were able to grow in soft agar<sup>9</sup> and caused tumour formation in nude mice<sup>10</sup> (Table 1). Untransfected NIH3T3 or Vero cells, or cells transfected with the pSV<sub>2</sub>neo plasmid only, were not transformed as judged by any of these parameters.

We tested for the presence of the yeast ATPase gene in the transfected cell lines and in some of the tumours induced in mice. Southern blot analysis<sup>11</sup> indicated the presence of one or two copies of a 4.2-kilobase (kb) *Hind*III fragment containing the yeast ATPase in the NIH and Vero cell lines which had been selected for positive expression with the ATPase antibody (Fig. 3a). Control cell lines did not reveal the presence of the ATPase gene. When the DNA from tumours induced by one ATPase-NIH3T3 clone in mice was analysed (Fig. 3b) we found that the ATPase gene was amplified by up to 15-fold (lanes 5 and 6) over the parental cell line (lane 7). This amplification is specific as the mouse *Mac-1* gene<sup>12</sup> used as a control showed no such increase (Fig. 3c). These findings suggest the tumours were produced from cells carrying the gene for the yeast ATPase, and that there was a positive selection for the cells having higher ATPase gene copy number.

Thus the gene for yeast  $\text{H}^+$ -ATPase appears to behave as an oncogene when expressed in animal cells. To find out if activity

of the yeast ATPase is required for cell transformation, we have used defective mutants constructed by directed mutagenesis. When expressed in yeast, the proton pumping activity of the Asp378→Glu and Lys474→Gln mutants were 20%, and less than 5% respectively, of the wild-type activity<sup>13</sup>. These mutant genes were cloned into the pSV<sub>2</sub> vector and cotransfected with pSV<sub>2</sub>neo as described above for the wild-type gene. After selection with G-418 we determined the percentage of clones with transformed morphology from the total number of G-418 resistant clones. Although the efficiency of transformation with the wild-type ATPase gene (plasmid pSV<sub>h</sub>ATP<sub>5</sub>) was 50%, the mutations Asp378→Glu and Lys474→Gln gave only 10% and less than 5% as the efficiency of transformation, respectively. No transformed foci were observed with the pSV<sub>2</sub>neo plasmid alone. In addition, clones expressing the low activity Lys474→Gln mutant ATPase (demonstrated by immunofluorescence) failed to produce tumours in nude mice under the conditions in Table 1. Therefore the oncogenic transformation of mammalian cells depends on the activity of the yeast ATPase.

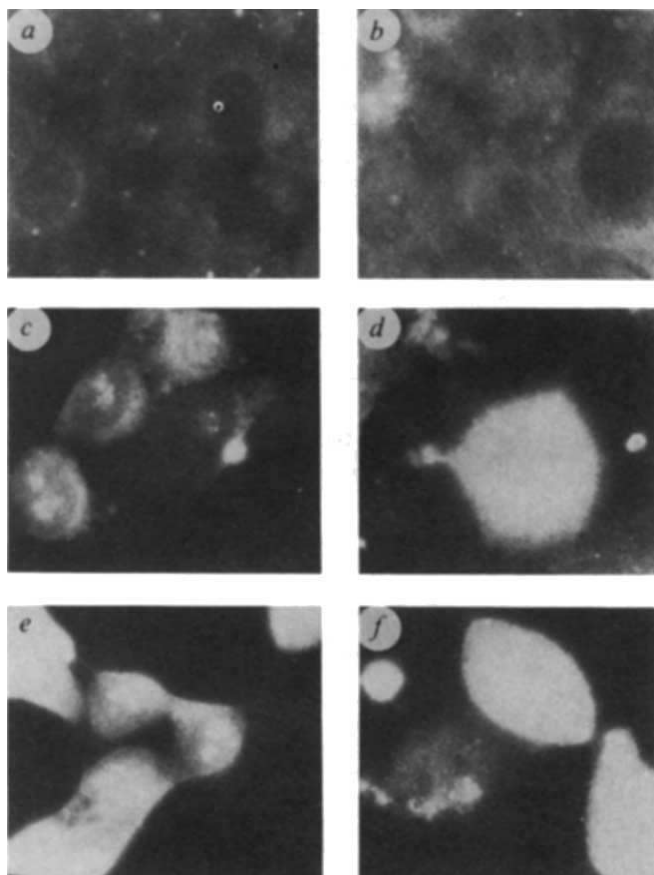
The intracellular pH of cells growing in the conditions of Table 1 was measured after removing the serum and incubating for 30 min in bicarbonate-free medium. The distribution of 5,5-dimethylthiazolidine-2,4-dione and benzoic acid was determined as described<sup>14</sup>, except that cell volume was also determined with an electronic cell sorter (Coulter Epics V). NIH3T3 cells expressing the yeast ATPase had a more alkaline pH than controls not expressing the ATPase or expressing the inactive Lys474→Gln mutant (pH 7.3 compared with 7.1, on an average of four determinations with a standard deviation of 0.06). Similar results were obtained with the Vero cell line, where the intracel-

Table 1 Properties of transfected cell lines

| Cell line           | Yeast ATPase | Morphology  | Saturation density | Soft agar cloning (%) efficiency) | Tumorigenicity in nude mice |         |
|---------------------|--------------|-------------|--------------------|-----------------------------------|-----------------------------|---------|
|                     |              |             |                    |                                   | Incidence                   | Latency |
| <i>NIH3T3 cells</i> |              |             |                    |                                   |                             |         |
| NneoI               | —            | Normal      | 1                  | <0.1                              | 0/5                         |         |
| NneoII              | —            | Normal      | 1                  | <0.1                              | 0/5                         |         |
| NneoIII             | —            | Normal      | 1                  | <0.1                              | 0/5                         |         |
| RNI1s               | +            | Transformed | 4                  | 20                                | 9/9                         | 7 days  |
| RNI3a               | +            | Transformed | 3                  | 10                                | 6/6                         | 4 days  |
| RNI1a               | +            | Transformed | 3                  | 12                                | 6/6                         | 4 days  |
| <i>Vero cells</i>   |              |             |                    |                                   |                             |         |
| VneoI               | —            | Normal      | 1                  | <0.1                              | 0/6                         |         |
| VneoII              | —            | Normal      | ND                 | <0.1                              | ND                          |         |
| VneoIII             | —            | Normal      | ND                 | <0.1                              | ND                          |         |
| RV13a               | +            | Transformed | 2                  | 3                                 | 2/6*                        | 26 days |
| RV13b1              | +            | Transformed | ND                 | 2                                 | ND                          |         |
| RV13c               | +            | Transformed | ND                 | 3                                 | ND                          |         |

Vero and NIH3T3 cells were seeded at a density of  $5 \times 10^5$  cells in Dulbecco's modified Eagle's medium supplemented with 10% calf serum. Transfections were performed<sup>8</sup> using 2  $\mu\text{g}$  pSV<sub>2</sub>neo plasmid, 8  $\mu\text{g}$  carrier DNA and (for the yeast ATPase expressing lines) 10  $\mu\text{g}$  pSV<sub>h</sub>ATP<sub>5</sub> plasmid (see text). After 3 days cells were treated with trypsin and aliquoted into 4 dishes. Transformant clones were selected with G-418. Stable clones were isolated by cylinder cloning and growth in mass culture. The yeast ATPase was detected by both immunofluorescence<sup>8</sup> and Southern analysis<sup>11</sup>. For plating in soft agar  $10^2$  and  $10^3$  cells were seeded into normal medium containing 0.25% agarose<sup>9</sup> and colonies counted after 14 days. Results are expressed as percentage of seeded cells. To test for *in vivo* growth potential<sup>10</sup> individual clones ( $10^6$  cells per animal) were washed with phosphate-buffered saline (without calcium) and injected subcutaneously into 30–40-day-old nude mice. The animals used for the Vero cell lines were irradiated (500 rads) 24 h before injection to eliminate natural killer cells. Results are expressed indicating the number of mice developing tumours/total number of animals and the time required for the appearance of the tumours where appropriate. For the saturation density assay the cells were seeded at  $10^5$  cells per 16 mm well, incubated in 2 ml medium (see above) for 6 days and counted. The results are expressed relative to control cells not expressing the yeast ATPase. ND, not determined.

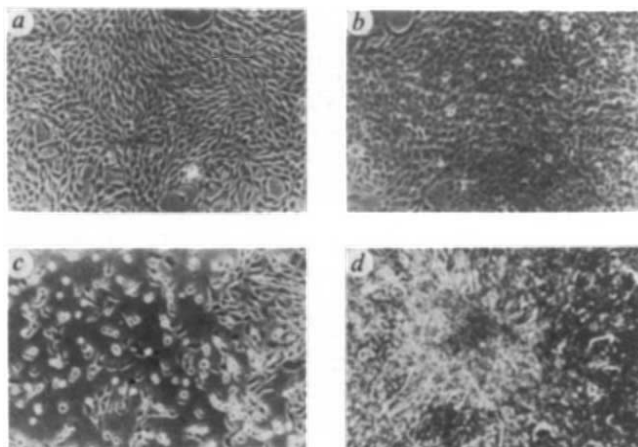
\* Tumours induced by the ATPase-Vero lines had a long latency period and reverted after about one week. This is probably due to the type of cells (monkey) because in our hands most of the tumours developed by HeLa cells (human) in nude mice also reverted after a short time.



**Fig. 1** Stable expression of yeast  $H^+$ -ATPase in NIH3T3 (left) and Vero (right) cells. Cells transfected with pSV<sub>2</sub>neo alone (a and b) or with both pSV<sub>2</sub>neo and pSV<sub>n</sub>AT<sub>5</sub> (c-f) were grown on cover slides, fixed with formaldehyde, permeabilized with Triton X-100 and incubated with a rabbit polyclonal antiserum against the yeast ATPase<sup>7</sup>. Cells were labelled with second antibody coupled to fluorescein isothiocyanate and visualized with a fluorescence microscope<sup>8</sup>. Micrographs are shown for the clones Nneol (a), Vneol (b), RN1ds (c), RVt3a (d), RN1a (e) and RVt3b1 (f) of Table 1.

lular pH of the control cells was 7.0, and that of ATPase-expressing cells was 7.2 (on an average of three determinations, with a standard deviation of 0.05).

We cannot discard the possibility that an unsuspected interaction of the yeast ATPase with mammalian cells resulted in transformation and the increase in intracellular pH by unknown mechanisms. However, the most plausible interpretation of our results is that a heterologous proton pump activates the proliferation of animal cells by continually extruding protons and causing a sustained alkalization of the cytoplasm. This implies that the increase in cytoplasmic pH may have a more active role in cell proliferation than as a 'permissive' factor in the initiation of DNA synthesis in growth-factor-mediated and oncogene-induced cell proliferation<sup>3-5</sup>. We emphasize that it was the expression of the yeast proton-pumping ATPase in animal cells that allowed the testing of the effect of sustained cellular alkalization on cell proliferation for the first time. As discussed in the introduction, previous methods used to increase intracellular pH suffered from long-term toxicity effects<sup>2</sup>. Although our results suggest that intracellular alkalization may trigger cell proliferation, a rigorous demonstration of a crucial role for pH increase will require detailed studies on the initial stages of cell transformation. We are approaching this problem by expressing the yeast ATPase under the control of



**Fig. 2** Morphology in phase micrographs ( $\times 75$ ) of representative transfected cell clones. G-418-resistant colonies generated by transfection of NIH3T3 (a and c) and Vero (b and d) cells with either pSV<sub>2</sub>neo alone (a and b) or with both pSV<sub>2</sub>neo and pSV<sub>n</sub>AT<sub>5</sub> resulting in expression of the yeast ATPase (b and d). Micrographs are shown for clones Nneol (a), Vneol (b), RN1ds (c) and RVt3a (d) of Table 1.

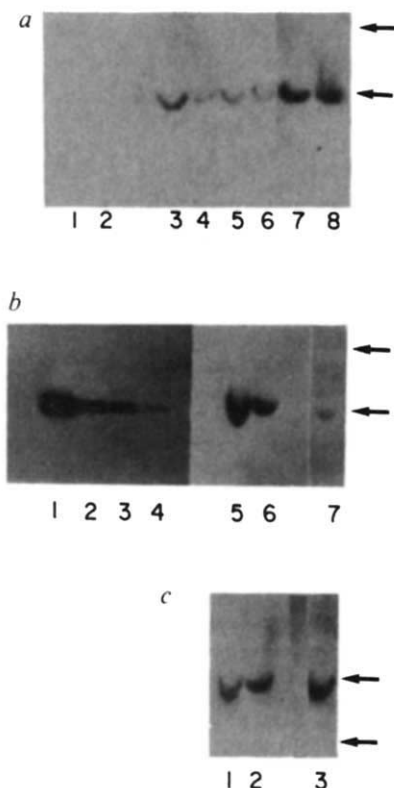
an adjustable promoter and by using temperature-sensitive mutants of the yeast ATPase gene.

It has been reported that microinjection of Ha-ras protein<sup>15</sup> and expression of *v-mos* and Ha-ras oncogenes<sup>16</sup> in quiescent NIH3T3 cells cause intracellular alkalization by activating the  $H^+/Na^+$  exchanger and this seems to induce cell-cycle progression. In addition, tumorigenic fibroblasts obtained by chemical mutagenesis have an activated  $H^+/Na^+$  exchanger and increased intracellular pH<sup>17</sup>. These experiments did not demonstrate that sustained alkalization was the primary transforming event but, as in our case, this seems most likely. Therefore, continuous activation of proton efflux in animal cells by different mechanisms may result in tumorigenic transformation. The regulatory mechanisms involved in the effect of pH on proliferation are not known. An increase in intracellular pH is known to activate glycolysis, protein and DNA synthesis and microtubule assembly<sup>2</sup>. In addition, specific effects on regulatory proteins controlling cell growth could also be considered. A genetic approach using yeast as model system may illuminate the nature of the pH-sensor(s) of eukaryotic cells. We are currently isolating mutations which suppress the growth defect of yeast strains with lowered proton-pumping activity and hope that some of them will be related to the pH-sensing mechanism controlling growth.

We have started to characterize the NIH3T3 lines expressing the ATPase and preliminary results indicate that they have lost the serum requirements for DNA synthesis. Future studies should also clarify the state of other signals connected with cell proliferation, such as cAMP<sup>18</sup>, calcium<sup>18</sup>, phosphoinositide lipid turnover<sup>19</sup> and tyrosine kinases<sup>20</sup>. It would also be interesting to explore the capability of the yeast ATPase gene to transform primary animal cells, in addition to the established fibroblast lines discussed here. It is also possible that tumours may be induced by oncogenes related to the proton transport systems of animal cell, such as the  $H^+/Na^+$  exchanger<sup>2-5</sup>, ( $H^+ + K^+$ ) ATPase<sup>21</sup> and vacuolar  $H^+$ -ATPase<sup>22</sup>. Finally, as the proliferation of plant cells also seems to be controlled by their proton-pumping activity<sup>6</sup>, the generation of hormone-independent calluses by expressing the yeast ATPase is a very attractive possibility.

This work was supported by a grant from the Spanish Comisión Asesora de Investigación Científica y Técnica. R.P. was a fellow of the Becas de Reincorporación programme of the Ministerio de Educación y Ciencia. We thank Dr Francisco Vara for the measurements of dimethylloxazolidinedione uptake,





**Fig. 3** Detection of the yeast ATPase gene in cells transformed by plasmid pSV<sub>h</sub>AT<sub>5</sub>. High molecular weight DNA (20 µg) obtained from the different cell lines and tumours was digested with *Hind*III, subjected to electrophoresis, blotted and hybridized with a <sup>32</sup>P-labelled, nick-translated 4.2-kb ATPase sequence from plasmid pSV<sub>h</sub>AT<sub>5</sub>. *a*, Lanes 1 and 2 correspond to the control cell lines Vneol and Neol, respectively; lanes 3–8 correspond to the ATPase-expressing lines RVt3a, RVt3b1, RV1e3, RNt3a, RN1a and RN1ds in the indicated order (see Table 1). *b*, Lanes 1–4 contain pSV<sub>h</sub>AT<sub>5</sub> plasmid digested with *Hind*III in amounts equivalent to 50, 10, 5 and 1 gene copies per genome, respectively; the other lanes contain DNA from tumours induced in Swiss (lane 5) and nude (lane 6) mice by cell line RN1ds, its DNA being analysed in lane 7. *c*, Same blot as above, washed and rehybridized with a 1.5-kb *Eco*RI fragment of the Mac-1 mouse gene<sup>12</sup>. The upper and lower arrows correspond to size markers of 23 and 4.4 kb, respectively.

Dr Francisco Portillo for ATPase mutants, Dr Thomas Graf for the NIH3T3 cell line and Dr Leandro Sastre for the Mac-1 probe. We also thank Drs Thomas Graf, Lennart Philipson, Angel Pestana and Leandro Sastre for helpful discussions. We are indebted to Dra. Carmina Criado and Carmen Palomo for the animal care assistance and Antonio Fernandez for the photographs.

Received 16 February; accepted 27 June 1988.

1. Nuccitelli, R. & Heiple, J. M. in *Intracellular pH: Its Measurement Regulation and Utilization in Cellular Functions* (eds Nuccitelli, R. & Deamer, D. W.) 567–586 (A. Liss, New York, 1982).
2. Epel, D. & Dube, F. in *Control of Animal Cell Proliferation* Vol. 2 (eds Boynton, A. L. & Leffert, H. L.) 363–393 (Academic, New York, 1987).
3. Rozengurt, E. in *Molecular Mechanisms of Transmembrane Signalling* (eds Cohen, P. & Houslay, M. D.) 429–452 (Elsevier, Amsterdam, 1985).
4. Pouyssegur, J., Franchi, A., Kohn, M., L'Allemain, G. & Paris, S. *Curr. Top. Membranes Transp.* **26**, 201–220 (1985).
5. Moolenaar, W. H. *A. Rev. Physiol.* **48**, 363–367 (1986).
6. Serrano, R. *Plasma Membrane ATPase of Plants and Fungi* (CRC, Boca Raton, 1985).
7. Serrano, R., Kiehlbrandt, M. C. & Fink, G. R. *Nature* **319**, 689–693 (1986).
8. Gorman, C. in *DNA Cloning* Vol. 2 (ed. Glover, D. M.) 143–190 (IRL, Oxford, 1985).
9. Jacoby, W. B. & Pastan, I. H. *Meth. Enzym.* **58**, 152–164 (1979).
10. Blair, D. G., Cooper, C. S., Oskarsson, M. K., Eader, L. A. & Vande Woude, G. F. *Science* **218**, 1122–1125 (1982).

11. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
12. Sastre, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 5644–5648 (1986).
13. Portillo, F. & Serrano, R. *EMBO J.* **7**, 1793–1798 (1988).
14. Vara, F., Schneider, J. A. & Rozengurt, E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2324–2388 (1985).
15. Hagag, N., Lacal, J. C., Graber, M., Aaronson, S. & Viola, M. V. *Molec. cell. Biol.* **7**, 1984–1988 (1987).
16. Doppler, W., Jaggi, R. & Groner, B. *Gene* **54**, 147–153 (1987).
17. Ober, S. S. & Pardee, A. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2766–2770 (1987).
18. Whitfield, J. F., Boynton, A. L., Rixon, R. H. & Yondale, T. in *Control of Animal Cell Proliferation* Vol. 1 (eds Boynton, A. L. & Leffert, H. L.) 331–365 (Academic, New York, 1985).
19. Downes, C. P. & Michel, R. H. in *Molecular Mechanisms of Transmembrane Signalling* (eds Cohen, P. & Houslay, M. D.) 3–56 (Elsevier, Amsterdam, 1985).
20. Foulkes, J. G. & Rosner, M. R. in *Molecular Mechanisms of Transmembrane Signalling* (eds Cohen, P. & Houslay, M. D.) 217–252 (Elsevier, Amsterdam, 1985).
21. Sachs, G., Faller, L. D. & Rabon, E. *J. Membrane Biol.* **64**, 123–144 (1982).
22. Mellman, I., Fuchs, R. & Helenius, A. *A. Rev. Biochem.* **55**, 663–700 (1986).

## Extensive variation of human immunodeficiency virus type-1 *in vivo*

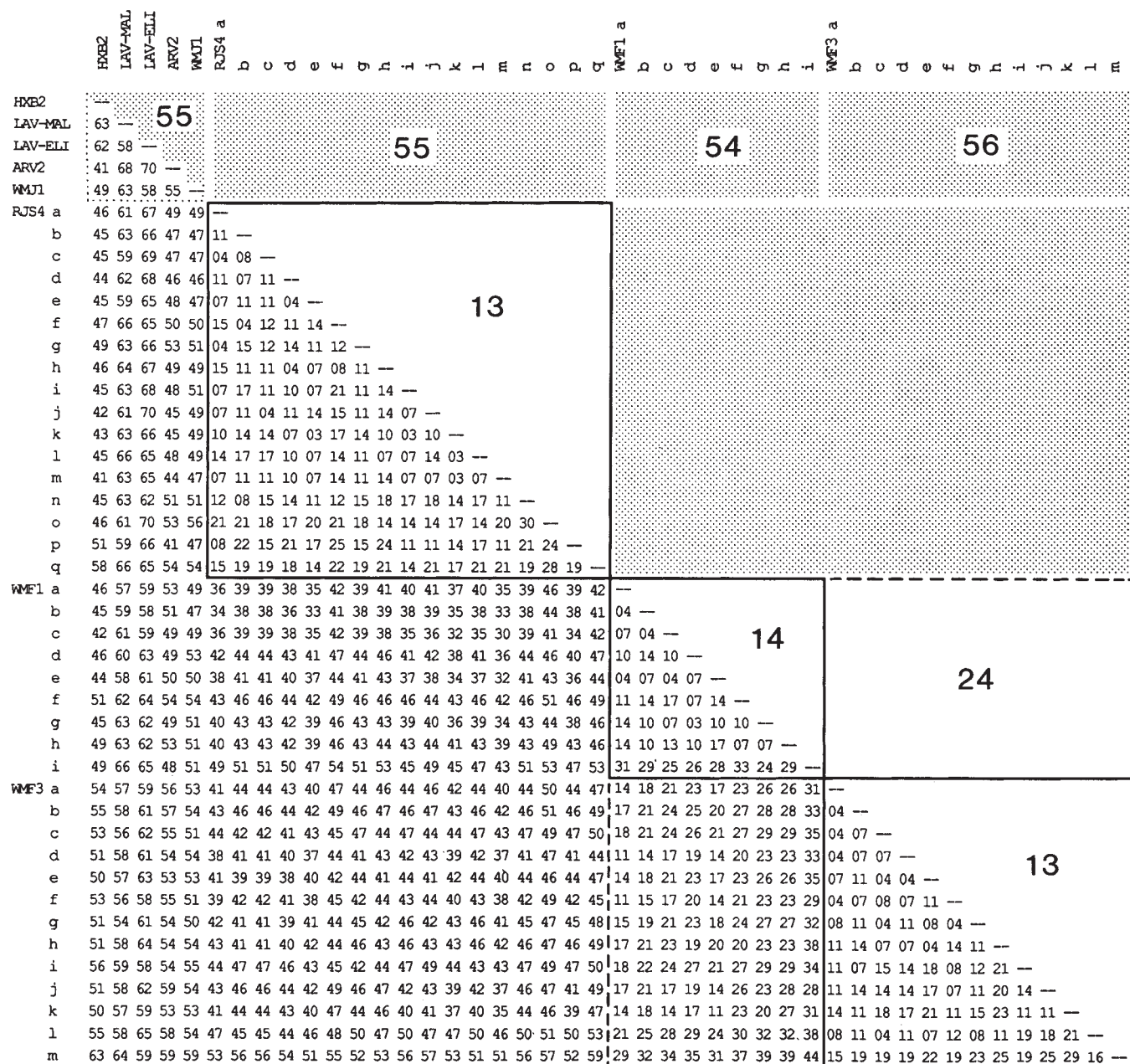
Michael S. Saag\*, Beatrice H. Hahn\*†, Joseph Gibbons\*, Yuexia Li\*†, Elizabeth S. Parks‡, Wade P. Parks‡, & George M. Shaw\*§||

Departments of Internal Medicine\*, Microbiology† and Biochemistry‡, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA  
Department of Pediatrics‡, University of Miami School of Medicine, Miami, Florida 33101, USA

Genotypic variation among independent isolates of human immunodeficiency virus type-1 (HIV-1) is well known<sup>1–9</sup>, but its molecular basis and biological consequences are poorly understood. We examined the genesis of molecular variation in HIV-1 by sequential virus isolations from two chronically infected individuals and analysis of recombinant HIV-1 genomic clones. In three different virus isolates full-length HIV-1 clones were identified and found to consist, respectively, of 17, 9 and 13 distinguishable, but highly-related, viral genotypes. Thirty-five viral clones derived from two HIV-1 isolates obtained from the same individual but 16 months apart showed progressive change, yet were clearly related. Similar changes in the HIV-1 genome did not occur *in vitro* during virus isolation and amplification. The results indicate that HIV-1 variation *in vivo* is rapid, that a remarkably large number of related but distinguishable genotypic variants evolve in parallel and coexist during chronic infection, and that 'isolates' of HIV-1, unless molecularly or biologically cloned, generally consist of complex mixtures of genotypically distinguishable viruses.

In previous studies, we observed that independent as well as sequential HIV-1 isolates from individual patients showed evidence of genetic change<sup>1,2,5,8</sup>. Genotypic analysis of paired virus isolates from donor-recipient transfusion cases<sup>10</sup>, and nucleotide sequence comparisons of geographically distinct HIV-1 strains<sup>4</sup>, provided further evidence for the highly variable nature of HIV-1. The extent and patterns of HIV-1 variation in chronically infected individuals, however, were not known. To address this question, we examined sequential virus isolates from two individuals (R.J.S. and W.M.F.)<sup>8</sup> at a molecular level. Recombinant lambda phage libraries were prepared from isolates RJS4, WMF1 and WMF3 so that individual complete viral DNA molecules, which together comprise the overall isolate DNA<sup>8</sup>, could be analysed. A total of 27, 17 and 18 full-length HIV-1 clones were obtained from the three libraries and detailed mapping using 11 restriction endonucleases revealed that 17 of 27 RJS4 clones, 9 of 17 WMF1 clones, and 13 of 18 WMF3

|| To whom correspondence should be addressed.



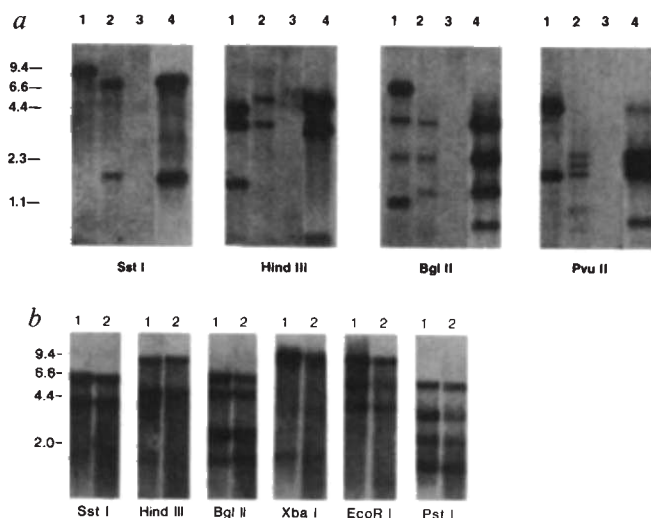
**Fig. 1** Analysis of variation among HIV-1 DNA clones. Each of the 60 clones was compared to every other clone pattern, and the percentage difference in restriction site differences between pairs of clones was calculated as follows:  $\frac{A + B - C}{2} \times 100 = \%$ , where, A equals the number of restriction sites present in clone X, B equals the number of restriction sites present in clone Y that are not present in clones X and Y combined, with identical sites in the pair counted once; and C equals the number of additional sites, they have: (0+0)/25×100=0% difference. If two sites but none are in common, they have: (25+25)/50×100=100% difference. To clone RJS and WMF, these enzyme sites were not included in the analysis. The *SstI* sites in all clones except WMF1b, c, g, h, and i actually represent two sites was included in the analysis of variation. Boxed in solid line indicate clones from the same isolate. Dashed lines highlight the comparison of viral isolates independent (unrelated) clones. Mean differences among clones from each isolate were calculated from raw data.

47 14 18 14 17 11 23 20 27 31 14 11 18 17 21 11 15 23 11 11 —  
53 21 25 28 29 24 30 32 32.38 08 11 04 11 07 12 08 11 19 18 21 —  
59 29 32 34 35 31 37 39 39 44 15 19 19 19 22 19 23 25 19 25 29 16 —

indistinguishable clone patterns depicted in Fig. 3: was compared pairwise  
endonuclease cleavage sites was calculated as follows:  $(A + B) / C \times 100 =$   
sites present in one clone (X) that are missing in the other clone (Y);  
missing in clone X; C equals the total number of restriction sites present  
only once. For example, if clone X and clone Y have 25 restriction sites  
restriction site differences. If clones X and Y each have 25 restriction  
restriction site differences. Because *EcoRI* and *SstI* enzymes were used  
analysis of variation among these clones. Also, because the two terminal  
redundant sequences in their long terminal repeats, only one of the  
are the difference scores for viral clones derived from the RJS4, WMF1  
clones between isolates WMF1 and WMF3, and dotted lines highlight  
within individual isolates (RJS4a-q; WMF1a-i; WMF3a-m) compared  
values, not percentages, and are shown in the open and shaded areas.

The genomic restriction patterns (Fig. 3) showed that clones  
from individual virus isolates (RJS4a-q; WMF1a-i; WMF3a-  
m), despite their heterogeneity, were considerably more similar  
to other clones within the same isolate than to unrelated clones.  
To quantify this variation, we carried out a pairwise comparison  
of the percentage of restriction site differences between each of  
the 44 viral genotypes shown in Fig. 3 (total of 946 independent  
comparisons; Fig. 1). Such an analysis is a valid means for  
estimating overall nucleotide sequence variability<sup>11,12</sup>; assuming





**Fig. 2** Southern blot hybridization patterns of HIV-1 DNA comparing (a) uncultured versus cultured brain tissue specimens and (b) co-cultures of lymphocytes from an HIV-1 infected patient with normal lymphocytes from two HLA-unrelated donors (lanes 1 and 2). In a, lanes 2 and 4 contain DNA from uncultured and cultured brain tissue of an AIDS patient, lane 1 DNA from an unrelated HIV-1 isolate, and lane 3 normal human lymphocyte DNA.

**Methods a**, Brain tissue was obtained from an AIDS patient at autopsy and divided. Half was minced and co-cultured with interleukin-2 stimulated normal donor lymphocytes for four weeks to isolate and amplify HIV-1, and the other half was extracted immediately for high molecular weight DNA without interim culture. **b**, Identical aliquots of PBL from a different HIV-1 infected subject were co-cultured for four weeks with PBL from two HLA-unrelated donors. DNA extractions and blot hybridizations with full-length HIV-1 probe (pBH10i) were performed as described<sup>1</sup>.

that the loss or gain of a restriction site in otherwise highly-related genomes results from a change in a single nucleotide, a 50% difference in restriction sites for enzymes that recognize six base pair sequences corresponds to about an 8% nucleotide sequence divergence (that is, 1 nucleotide change out of 12 nucleotides sampled). We used this approach first to analyse clones of HIV-1 for which nucleotide sequence information was available in the GenBank database, Los Alamos National Laboratory, and the correlation between predicted and observed nucleotide sequence differences was found to be quite good. For example, clone HXB2 differs from clones BH10, ARV2 and LAVeli respectively by 13%, 41% and 62% in restriction sites (ref. 1 and Fig. 1). Thus the predicted nucleotide differences are 2.2%, 6.8% and 10.3% and the actual differences determined by nucleotide sequence comparisons are 1.6%, 5.8% and 9.7%. Similarly, clones LAVmal, LAVeli and ARV2 (Fig. 1) differ by 58–70% in restriction sites, by 10–12% in predicted nucleotide sequence and by 10.1–13.0% in sequence determined experimentally (and deposited in GenBank).

The restriction site differences among viral DNA genomes comprising the individual isolates RJS4, WMF1 and WMF3 were considerably less than those of viral genomes representing unrelated (independent) isolates. For example, the 17 different RJS4 genotypes varied from each other by 3–28% (mean of 13%) whereas the same clones varied from independently isolated viruses by 41–70% (mean of 55%;  $P < 0.0001$ , chi-square analysis). The nine different WMF1 genotypes varied from each other by 3–33% (mean of 14%) and from independent viruses by 42–66% (mean of 54%;  $P < 0.0001$ ). The 13 different WMF3 genotypes varied from each other by 4–29% (mean of 13%) and from independent viruses by 50–65% (mean of 56%;  $P < 0.0001$ ).

Isolates WMF1 and WMF3 were derived from cultures of peripheral blood lymphocytes of the same individual taken 16 months apart. If the viral forms present in WMF3 had evolved from viruses present in WMF1, or from common precursor viruses in that patient, the extent of similarity between WMF1 and WMF3 viral genomes should be intermediate between values for clones from within each isolate, and clones from unrelated isolates. Conversely, if superinfection with unrelated viruses had occurred, genotypes of WMF1 and WMF3 viruses would be expected to vary to the same extent as the unrelated isolates HXB2, LAVmal, LAVeli, ARV2 and WMJ1. The data in Fig. 1 show that the viral clones of WMF1 (a–i) differed from clones of WMF3 (a–m) by 11–44% (mean of 24%), a range intermediate between values for viruses from within single isolates (mean of 13%;  $P < 0.0001$ ) and from various unrelated isolates (mean of 55%;  $P < 0.0001$ ). These data indicate that the viruses comprising WMF3 evolved either from genomes present in WMF1, or from precursor genomes common to both. The absence of clones in WMF3 that are identical to clones in WMF1, the existence of viral genotypes in WMF1 and WMF3 that differ in restriction pattern by 11–44% (2–7% estimated nucleotide sequence difference), and the fact that most viral species comprising an 'isolate' actually differ from the most prevalent genotypic form within the isolate (for example, RJS4a versus RJS4b–q), underscore the extremely rapid rate of change of HIV-1 *in vivo*.

As a control for these studies, a recombinant lambda library identical to those prepared for RJS4, WMF1 and WMF3 was made from an isolate of HIV-1 that had been biologically-cloned by five sequential cell-free end-point dilutions of culture supernatant. This was done to obtain a genotypically-pure virus stock that could then be expanded in culture (as is done during the isolation of HIV-1 from human tissues) to determine how much genotypic variation is introduced during virus amplification *in vitro*. From this library, ten full-length viral DNA clones were identified and found to be identical in 270 out of 273 restriction sites mapped (average of 27 sites per clone; data not shown). This minimal degree of restriction site variability in virus forced through multiple rounds of *in vitro* replication was much less than that of the isolates RJS4, WMF1 and WMF3, and could

**Fig. 3** Restriction endonuclease cleavage patterns for seven independent HIV-1 viral clones (HXB2; LAVmal; LAVeli; ARV2; WMJ1; RJS4a; WMF1a) and for distinguishable viral clones from isolates RJS4, WMF1 and WMF3 (ref. 8). Shown to the right of each set of maps is a summary of the different restriction patterns (E, *EcoRI*; S, *SstI*; X, *XhoI*; B, *BglII*; U, *PvuII*; H, *HindIII*; P, *PstI*; K, *KpnI*; M, *BamHI*; C, *CvnI*; A, *XbaI*) with the top clone pattern designated 1111111111 and differences in subsequent clone patterns identified sequentially. The laboratory clone designation (for example, RJS4.1; RJS4.11; RJS4.16) and the total number of viral clones represented by each pattern are indicated.

**Methods.** Isolates RJS4, WMF1 and WMF3 were derived from peripheral blood lymphocytes (PBL) of HIV-1 infected individuals as previously described<sup>8</sup>. Non-amplified recombinant phage libraries were prepared in  $\lambda$ -gt10 and screened with [<sup>32</sup>P]BH10i (a full-length HIV-1 probe) by standard techniques<sup>1</sup>. RJS4 was cloned in permuted form by digestion of circular DNA intermediate forms using *EcoRI*. WMF1 and WMF3 were cloned from integrated and unintegrated viral DNA in non-permuted form using *SstI*. All 62 full-length HIV-1 clones that were identified were mapped side by side using Southern blot hybridization, single, double and triple enzyme digests, and full-length, 5' and 3' HIV-1 probes as described<sup>1</sup>. Clones comprising WMF1b, c, g, h and i patterns are ~170 base pairs shorter than the other clones due to a polymorphic *SstI* site located in the 5' leader sequence<sup>1</sup>. The restriction maps of HXB2, LAVmal, LAVeli, ARV2 and WMJ1 were obtained from their nucleotide sequences (GenBank) or from previous publications<sup>8</sup>.

|        |                |                 |         |      |         |           |
|--------|----------------|-----------------|---------|------|---------|-----------|
| HXB2   | SHS HU P H B C | U K K U E       | E SCH K | BU C | B B H M | XCKUBUS   |
| LAVmal | SH A U P       | CP MK K UAPE    | M C K U | B C  | BA C    | CK U S    |
| LAVeli | SHS HU P H     | P C K K PE      | M M P U | CE   | B       | X KABU S  |
| ARV2   | SHSPHUP P H B  | U K K UA H      | E ASCH  | BU C | C B P   | XCK UBS   |
| WMJ1   | SHS HU P U U   | P E C K K U E H | K       | B C  | B C H   | P XCK UBS |
| RJS4a  | SHS U P        | P UC K K U H    | E SP K  | BU C | A MP    | CKU UBS   |
| WMF1a  | SH HU P H      | C UC K K U      | EM P K  | B C  | AH P    | XCKU UBS  |

| Virus  | Pattern               |
|--------|-----------------------|
|        | E S X B U H P K M C A |
| HXB2   | 1 1 1 1 1 1 1 1 1 1   |
| LAVmal | 2 2 2 2 2 2 1 2 2 2   |
| LAVeli | 3 3 1 3 3 3 2 3 3 3   |
| ARV2   | 4 1 1 4 4 4 4 2 4 4 4 |
| WMJ1   | 5 3 1 5 5 5 1 4 5 1   |
| RJS4a  | 4 1 2 6 1 6 6 1 1 6 5 |
| WMF1a  | 4 2 1 6 6 7 7 1 5 7 5 |

|       |          |                  |        |      |      |         |
|-------|----------|------------------|--------|------|------|---------|
| RJS4a | SHS U P  | P UC K K U H     | E SP K | BU C | A MP | CKU UBS |
| b     | SHS HU P | P UC K K H       | E S K  | BU C | A MP | CKU UBS |
| c     | SHS U P  | P UC K K U H     | E S K  | BU C | A MP | CKU UBS |
| d     | SHS HU P | K P UC K K U H   | E S K  | BU C | A MP | CKU UBS |
| e     | SHS HU P | K P UC K K U H   | E SP K | BU C | A MP | CKU UBS |
| f     | SHS HU P | P UC K K H       | E S K  | BU C | A MP | KU UBS  |
| g     | SHS U P  | P UC K K U H     | E SP K | BU C | A MP | KU UBS  |
| h     | SHS HU P | K P UC K K U H   | E S K  | BU C | A MP | KU UBS  |
| i     | SHS U P  | B K P UC K K U H | E SP K | BU C | A MP | CKU UBS |
| j     | SHS U P  | B P UC K K U H   | E S K  | BU C | A MP | CKU UBS |
| k     | SHS HU P | B K P UC K K U H | E SP K | BU C | A MP | CKU UBS |
| l     | SHS HU P | B K P UC K K U H | E SP K | BU C | A MP | KU UBS  |
| m     | SHS HU P | B P UC K K U H   | E SP K | BU C | A MP | CKU UBS |
| n     | SHS HU P | P UC K K H       | E SP K | BU C | A M  | CKU UBS |
| o     | SHSA U P | B K P UC K K U   | E S K  | BU C | A MP | KU UBS  |
| p     | SHS U P  | B P UC K K U H   | E SP K | BU C | A P  | CK UBS  |
| q     | SHS U P  | K P UC K K H     | E SP K | U C  | A P  | CKU UBS |

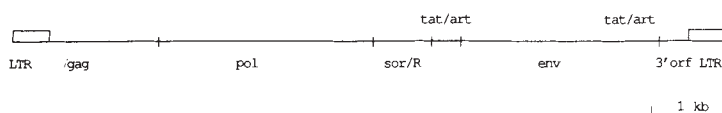
| Virus | Pattern               | Clone             | Total Number |
|-------|-----------------------|-------------------|--------------|
|       | E S X B U H P K M C A |                   |              |
| RJS4  |                       |                   |              |
| a     | 1 1 1 1 1 1 1 1 1 1   | 1, 11, 16, 39, 46 | 5            |
| b     | 1 1 1 1 2 2 2 1 1 1   | 2, 23, 26         | 3            |
| c     | 1 1 1 1 1 2 1 1 1 1   | 18, 34            | 2            |
| d     | 1 1 1 1 1 2 2 1 1 1   | 15, 40            | 2            |
| e     | 1 1 1 1 1 2 1 2 1 1   | 20, 24            | 2            |
| f     | 1 1 1 1 2 2 2 1 1 2   | 12, 31            | 2            |
| g     | 1 1 1 1 1 1 1 1 2 1   | 7                 | 1            |
| h     | 1 1 1 1 1 2 2 2 1 2   | 32                | 1            |
| i     | 1 1 1 2 1 1 1 2 1 1   | 42                | 1            |
| j     | 1 1 1 2 1 1 2 1 1 1   | 6                 | 1            |
| k     | 1 1 1 2 1 2 1 2 1 1   | 22                | 1            |
| l     | 1 1 1 2 1 2 1 2 1 2   | 14                | 1            |
| m     | 1 1 1 2 1 2 1 1 1 1   | 21                | 1            |
| n     | 1 1 1 1 2 2 3 1 1 1   | 45                | 1            |
| o     | 1 1 1 2 1 3 2 2 1 2   | 19                | 1            |
| p     | 1 1 1 2 3 1 1 1 2 1   | 43                | 1            |
| q     | 1 1 1 3 2 1 1 2 2 1   | 4                 | 1            |

|       |             |             |           |     |        |          |
|-------|-------------|-------------|-----------|-----|--------|----------|
| WMF1a | SH HU P H   | C UC K K U  | EM P K    | B C | AH P   | XCKU UBS |
| b     | S HU P H    | C UC K K U  | EM P K    | B C | AH P   | XCKU UBS |
| c     | S HU P H B  | C UC K K U  | EM P K    | B C | AH P   | XCKU UBS |
| d     | SH HU P H B | C UC K K U  | EMAX P K  | B C | AH P   | XCKU UBS |
| e     | SH HU P H B | C UC K K U  | EM P K    | B C | AH P   | XCKU UBS |
| f     | SH HU H     | C UC K K U  | EMAX P K  | B C | AH P   | XCKU UBS |
| g     | S HU P H B  | C UC K K U  | EMAX P K  | B C | AH P   | XCKU UBS |
| h     | S HU P H    | C UC BK K U | EMAX P K  | B C | AH P   | XCKU UBS |
| i     | S HU PA H B | B K K U     | EMA P K C | B C | C AH P | XCKU UBS |

| Virus | Pattern               | Clone               | Total Number |
|-------|-----------------------|---------------------|--------------|
|       | E S X B U H P K M C A |                     |              |
| WMF1  |                       |                     |              |
| a     | 1 1 1 1 1 1 1 1 1 1   | 4, 5, 9, 15, 16, 18 | 6            |
| b     | 1 2 1 1 1 1 1 1 1 1   | 14, 19              | 2            |
| c     | 1 2 1 2 1 1 1 1 1 1   | 11, 13              | 2            |
| d     | 1 1 2 2 1 1 1 1 1 2   | 1, 12               | 2            |
| e     | 1 1 1 2 1 1 1 1 1 1   | 10                  | 1            |
| f     | 1 1 2 1 1 1 2 1 1 2   | 2                   | 1            |
| g     | 1 2 2 2 1 1 1 1 1 2   | 6                   | 1            |
| h     | 1 2 2 3 1 1 1 1 1 2   | 3                   | 1            |
| i     | 1 2 1 4 2 1 1 1 1 2 3 | 17                  | 1            |

|       |             |             |          |       |           |          |
|-------|-------------|-------------|----------|-------|-----------|----------|
| WMF3a | SH HU P BH  | BC C K K U  | EM P K   | B C   | A P       | XCKU UBS |
| b     | SH HU P BH  | BC C K K U  | EM P K   | B C K | A P       | XCKU UBS |
| c     | SH HU P BH  | BC C K K U  | EM K     | B C   | A P       | XCKU UBS |
| d     | SH HU P BH  | BC UC K K U | EM P K   | B C   | A P       | XCKU UBS |
| e     | SH HU P BH  | BC UC K K U | EM K     | B C   | A P       | XCKU UBS |
| f     | SH HU P H   | BC C K K U  | EM P K   | B C   | A P       | XCKU UBS |
| g     | SH HU P H   | BC C K K U  | EM K     | B C   | A P       | XCKU UBS |
| h     | SH HU P BH  | BC UC K K U | EM X K   | B C   | A P       | XCKU UBS |
| i     | SH HU P H   | BC C K K U  | EM P K   | B C K | A P X KU  | UBS      |
| j     | SH HU P H B | BC C K K U  | B EM P K | B C   | A P       | XCKU UBS |
| k     | SH HU P H B | C C K K U   | EM P K   | B C K | A P       | XCKU UBS |
| l     | SH HU P BH  | BC C K K U  | EM K     | B C   | A P X C U | UBS      |
| m     | SH HU P BH  | BC C K K U  | EM P     | B C   | A P X     | UBS      |

| Virus | Pattern               | Clone       | Total Number |
|-------|-----------------------|-------------|--------------|
|       | E S X B U H P K M C A |             |              |
| WMF3  |                       |             |              |
| a     | 1 1 1 1 1 1 1 1 1 1   | 1, 110, 111 | 3            |
| b     | 1 1 1 1 1 1 1 2 1 1   | 103, 107    | 2            |
| c     | 1 1 1 1 1 1 2 1 1 1   | 3, 108      | 2            |
| d     | 1 1 1 1 2 1 1 1 1 1   | 10, 104     | 2            |
| e     | 1 1 1 1 2 1 2 1 1 1   | 102         | 1            |
| f     | 1 1 1 2 1 1 1 1 1 1   | 11          | 1            |
| g     | 1 1 1 2 1 1 2 1 1 1   | 101         | 1            |
| h     | 1 1 2 1 2 1 2 1 1 1   | 106         | 1            |
| i     | 1 1 1 2 1 1 1 2 1 2   | 9           | 1            |
| j     | 1 1 1 3 1 1 1 1 1 1   | 7           | 1            |
| k     | 1 1 1 4 1 1 1 2 1 1   | 12          | 1            |
| l     | 1 1 1 1 1 1 2 3 1 1   | 105         | 1            |
| m     | 1 1 1 1 3 1 1 4 1 2   | 109         | 1            |





even have resulted from incomplete biological cloning of the virus stock before final amplification. In a second set of experiments (Fig. 2), viral DNA obtained from cultured versus uncultured tissue specimens, and from parallel co-cultures of patient P.B.L. with lymphocytes from different normal donors, were found to have identical genotypic patterns. These data suggest that virus populations isolated by short-term culture of patient tissues are generally representative of viral populations present *in vivo*, although it is well established that under conditions of selective pressure genetic changes in the HIV-1 genome can occur *in vitro*<sup>13</sup>.

Previously, we mapped and sequenced predominant viral clones representing three sequential HIV-1 isolates (WMJ1, WMJ2, WMJ3) from a child with AIDS and calculated the rate of viral evolution to be on the order of  $10^{-3}$  nucleotide substitutions per site per year<sup>8</sup>. The current study extends these findings by demonstrating rapid, parallel evolution of large numbers of related but distinguishable HIV-1 genotypes during chronic viral infection. These data are similar to those reported for equine infectious anaemia virus (EIAV) in which parallel evolution of multiple genotypic and antigenic variants has been documented in experimentally infected animals<sup>14,15</sup>. The potential molecular mechanisms underlying HIV-1 variation have been discussed<sup>4-9</sup>. It is of interest and potential clinical relevance that in none of the RJS, WMF or WMJ genomic libraries, nor in a large number of virus isolates from multiply exposed high-risk individuals<sup>1,2</sup>, was there evidence for concomitant superinfection with HIV-1 strains that were genotypically unrelated to the predominant viral forms. Whether this was due to selective pressures of *in vitro* cultivation or to *in vivo* mechanisms that protect against superinfection is unknown.

The biological and immunological significance of HIV-1 variation in viral pathogenesis is currently uncertain, but there are indications that similar variation in FeLV<sup>16</sup>, EIAV<sup>14,15,17</sup> and visna<sup>18,19</sup> is important. There are also reports that some isolates of HIV-1 have a preferential tropism for mononuclear phagocytes<sup>20,21</sup>, and, more recently, that more virulent forms of HIV-1 may become apparent as clinical immunodeficiency progresses<sup>22</sup>. Ten of 12 clones of RJS4, WMF1 and WMF3 that were tested for transfection competence gave virions that were morphologically indistinguishable from wild type, but which expressed biological phenotypes ranging from highly cytopathic lymphotropic viruses, to forms that replicated selectively in monocytes or not at all (A. Fisher, L. Kong and G. M. S., unpublished data). Future studies examining the genetic, biological and immunological properties of HIV-1 will need to account for the extensive variability of this virus.

We thank Jennifer Holt for editorial assistance, Gretchen Cloud for statistical analyses, and Maria Taylor and Rebecca Bunzel for technical assistance. This work was supported in part by grants from the NIH, the US Army Medical Research Acquisition Activity, and the Life and Health Insurance Medical Research Fund. M.S.S. is a Frommeyer Fellow in Investigative Medicine, B.H.H. a Special Fellow of the Leukemia Society of America, and G.M.S. a Pew Scholar in the Biomedical Sciences.

Received 7 March; accepted 2 June 1988.

- Shaw, G. M. *et al. Science* **226**, 1165-1171 (1984).
- Wong-Staal, F. *et al. Science* **229**, 759-762 (1985).
- Benn, S. *et al. Science* **230**, 949-951 (1985).
- Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. *Cell* **46**, 63-74 (1986).
- Starcich, B. *et al. Cell* **45**, 637-648 (1986).
- Willey, R. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 5038-5042 (1986).
- Desai, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 8380-8384 (1986).
- Hahn, B. *et al. Science* **232**, 1548-1553 (1986).
- Coffin, J. M. *Cell* **46**, 1-4 (1986).
- Srinivasan, A. *et al. Blood* **69**, 1766-1770 (1987).
- Desrosiers, R. C. *et al. J. Virol.* **54**, 552-560 (1985).
- Kestler, H. W. *et al. Nature* **331**, 619-621 (1988).
- Willey, R. L. *et al. J. Virol.* **62**, 139-147 (1987).
- Payne, S., Salinovich, O., Nauman, S., Issel, C. & Montelaro, R. *J. Virol.* **61**, 1266-1270 (1987).
- Payne, S. *et al. Virology* **161**, 321-331 (1987).
- Overbaugh, J., Donahue, P. R., Quackenbush, S. L., Hoover, E. A. & Mullins, J. I. *Science* **239**, 906-910 (1988).

- Montelaro, R. C., Parekh, B., Orrego, A. & Issel, C. J. *J. biol. Chem.* **259**, 10539-10544 (1984).
- Clements, J. E., Pedersen, F. S., Narayan, O. & Haseltine, W. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4454-4458 (1980).
- Narayan, O., Clements, J., Kennedy-Stoskopf, S., Sheffer, D. & Royal, W. *Contr. Microbiol. Immunol.* **8**, 60-76 (1987).
- Gartner, S., Markovits, P., Markovits, D. M., Betts, R. F. & Popovic, M. *Science* **233**, 215-219 (1986).
- Koyanagi, Y. *et al. Science* **236**, 819-822 (1987).
- Cheng-Meyer, C., Seto, D., Tateno, M. & Levy, J. A. *Science* **240**, 80-82 (1988).

## Biologically diverse molecular variants within a single HIV-1 isolate

Amanda G. Fisher\*†, Barbara Ensoli†, David Looney‡, Andrea Rose†, Robert C. Gallo†, Michael S. Saag§, George M. Shaw§, Beatrice H. Hahn§ & Flossie Wong-Staal†

† Laboratory of Tumour Cell Biology, NCI, NIH, Bethesda, Maryland 20892, USA

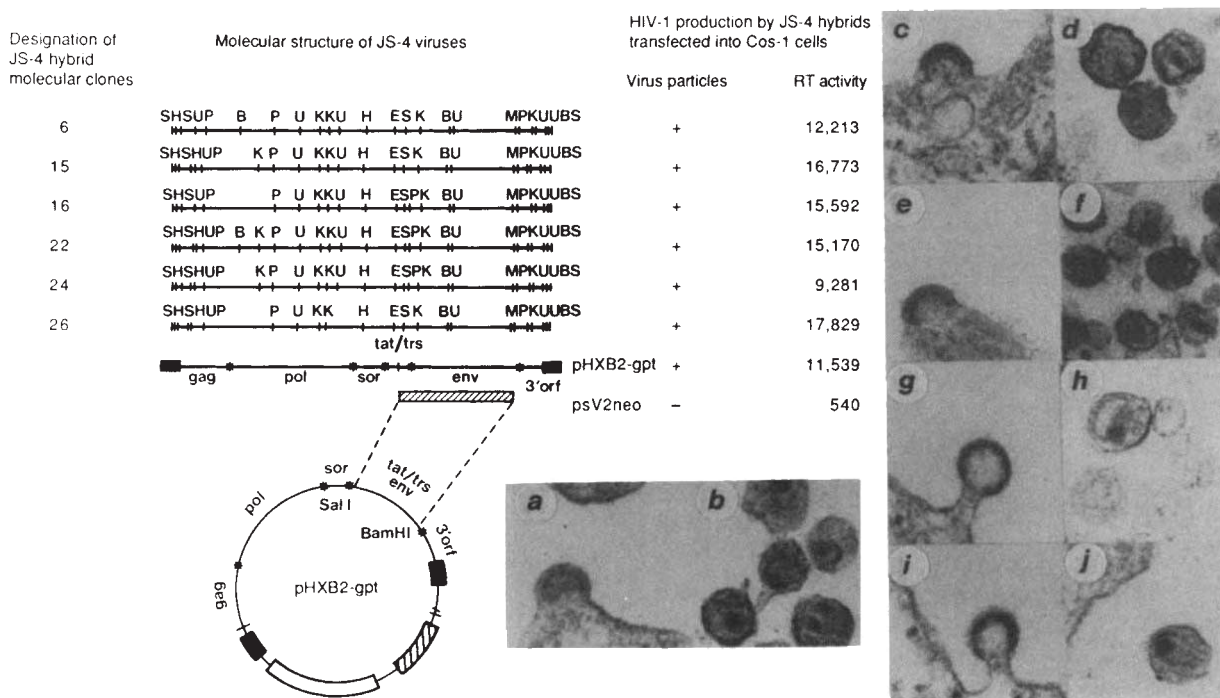
‡ Department of Viral Diseases, Walter Reed Army Institute of Research, Washington DC 20307, USA

§ Division of Haematology and Oncology, University of Alabama at Birmingham, University Station, Birmingham, Alabama 35295, USA

**AIDS is a disorder characterized by a slow progressive impairment of immune function and by infection of human immunodeficiency viruses (HIV-1, HIV-2)<sup>1-4</sup>. Our knowledge of how these viruses cause disease in man, or how the related lentiviruses (visna and equine infectious anaemia virus) cause disease in animals, is still fragmentary. In particular, the significance of genetic variation in HIV-1, occurring within populations, within individuals and over periods of time<sup>5,6</sup>, and the mechanisms of viral persistence remain unclear. To address these issues we prepared a series of proviral clones of HIV-1 originating from a single patient and compared their biological properties. Here we show that hybrid genomes (in which the envelope region of six viral clones were separately substituted into a prototype HIV-1 genome) generated viruses with widely differing capacity to grow in human T cells, cell lines and monocytoid cultures. These data suggest that extensive biological variation exists *in vivo* within an infected individual and is in part determined at the level of the viral envelope.**

Virus was isolated from the peripheral blood of an HIV-1 infected individual (coded R.J.S., HIV-1 isolate 4), who was a promiscuous homosexual male, chronically infected with HIV-1 for 5 years. Previous analyses had shown that serial isolates from R.J.S. were highly related and comprised a mixture of similar but distinct viral substrains<sup>7</sup>. Molecular clones were made of 27 full length proviral forms from isolate 4 and these were found to represent at least 17 different prototypes<sup>8</sup>. Of these, six clones were selected for further study (numbers 6, 15, 16, 22, 24 and 26) as detailed in Fig. 1. Because these proviruses were cloned using *EcoRI* (an enzyme which severs the viral genome)<sup>8</sup> and thus cloned in a permuted form, we isolated *env*-containing fragments (nucleotides 5,364-8,054) from each clone and placed these into the corresponding sites in the biologically active molecular clone pHXB2gpt<sup>9</sup>. This, we reasoned, would provide a panel of related viruses (designated JS4) in which the effects of envelope variation could be measured independently of the effects created by differences in other areas of the genome. It was also hoped that the constituents of such a panel would serve as useful reagents for neutralization studies. As shown in Fig. 1, each of the hybrid JS4 clones generated morphologically normal virus particles (Fig. 1c-1j) and high levels of reverse transcriptase upon transfection into the COS-1

\* Present address: ICRF Human Tumour Immunology Unit, Courtauld Building, 91 Riding House Street, London W1P 8BT, UK.



**Fig. 1** Construction and properties of JS4 hybrid viruses. Molecular clones of R.J.S.<sup>8</sup> were mapped by restriction enzyme digests using *Sst*I (S), *Eco*RI (E), *Bgl*II (B), *Pvu*II (U), *Hind*III (H), *Pst*I (P), *Kpn*I (K) and *Bam*HI (M) and used to prepare hybrid genomes in which the *env*-containing, *Sal*I-*Bam*HI fragment (2.7 kb) of each clone was substituted into corresponding sites in the proviral clone pHXB2gpt<sup>9</sup>. Solid boxes, crossed-hatched boxes and open boxes depict the HIV-1 long terminal repeats (LTRs), the bacterial gene for xanthine guanine phosphoribosyltransferase and a region containing the genes for ampicillin, SP6 and SV40, respectively. Electron micrographs of COS-1 cells following transfection (day 3) with hybrid clones show budding and mature HIV-1 virus in pHXB2gpt (a and b, respectively), JS4-15 (c and d, respectively), JS4-22 (e and f, respectively), JS4-16 (g and h, respectively), JS4-26 (i and j, respectively), JS4-6 and JS4-24 (not shown)-derived samples. Culture supernatants removed 5 days after transfection were concentrated 20-fold before performing assays for reverse transcriptase (RT). The results shown are the mean of three to seven independent experiments, except for a single value obtained for JS4-24. Values shown represent disintegrations per minutes (d.p.m.) of <sup>3</sup>H incorporated with the use of dT.rA as a primer template (values of <1,500 were obtained using dT.dA as a template).

**Methods.** Molecular analyses and subcloning were performed using conventional approaches<sup>16</sup>. Transfection was carried out using  $1 \times 10^6$  COS-1 cells and 10  $\mu$ g of plasmid DNA<sup>17</sup> and reverse transcriptase assays were performed using standard protocols<sup>18</sup>.

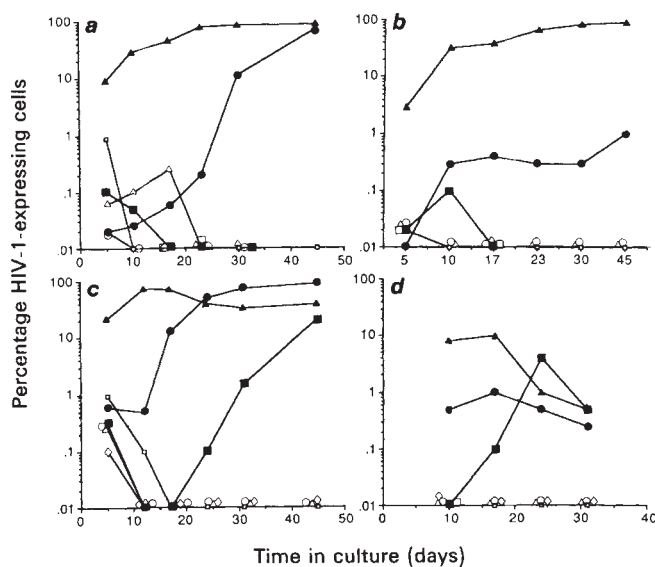
cell line. These data suggest that elements such as *gag*, *pol*, *tat* and *tr�/art*, which are critical for virus production, were functional in each of the JS4 clones.

The biological properties of these viruses were investigated in transmission studies in which transfected COS-1 cells were cocultured with various target cells and the incidence of HIV-1 expressing cells in the recipient populations was monitored with time (see Fig. 2). JS4 viruses proved extremely difficult to propagate in culture. In the early period of cocultivation (3–5 days) with transfected COS-1 cells, recipient H9 (Fig. 2a), MOLT-3 (Fig. 2b) and CEM cell cultures contained syncytial cells and rare HIV-1 *gag*-expressing cells (data not shown), but the frequency of expressing cells decreased with time until none were detected at 20–24 days in JS4-6, 15, 24 and 26 infected cultures. Productive infection of these target cell lines was achieved with JS4-22, but it required more than 30 and 45 days for 10% of H9 and MOLT-3 cultures respectively to be judged HIV-1 *gag*-p24 positive, whereas the parental HXB2 strain achieved this level within 7 days. Similar results were obtained using Jurkat target cells (Fig. 2c), with 10% infection being achieved by HXB2, JS4-22 and JS4-6 after <7, 15 and 37 days respectively, while JS4-15, 16 and 26 did not maintain productive infection. Using phytohaemagglutinin-activated cord blood T cells as targets (CB) (Fig. 2d) clones JS4-6 and 22 did establish productive infection, but at a lower and slower level than the parental clone HXB2. No HIV-1 expressing cells were seen in

cultures derived from clones JS4-15, 16, 24 and 26, and Southern blot analysis failed to detect viral DNA in these cell populations.

When cells, or virus-containing supernatants, were removed from cultures at the end of these experiments (day 31) and used to infect fresh CB cultures, the kinetics of infection in the secondary cultures paralleled those seen previously, suggesting that the infection rates are intrinsic properties of the viral strains. As all clones carry the same genetic information as HXB2, except for sequences contained within nucleotides 5,364–8,054, this region (which includes portions of *env*, *tat*, *tr�/art* and *orf-1* from the positive strand) must house the elements responsible for the altered phenotypes. We confirmed that the *tat* and *tr�* functions are preserved in the JS4 viruses by showing that each of the JS4 clones transactivated at levels comparable with that seen for HXB2 and that the level and pattern of viral mRNA expression in COS-1 cells transfected with JS4 clones were indistinguishable from those of the wild-type clone (data not shown). These data do not, of course, exclude subtle alterations in these genes that might contribute to the biological properties of the viruses. It is also conceivable that JS4 viruses have altered phenotypes because regulatory elements (such as TAR) present in the native R.J.S. proviruses have been removed in forming the hybrid genomes, and so we performed additional transfections using concatamerized DNA from the parental, full-length R.J.S. clones as controls. Preliminary results (data not shown) indicate that the biological properties of matched parental and



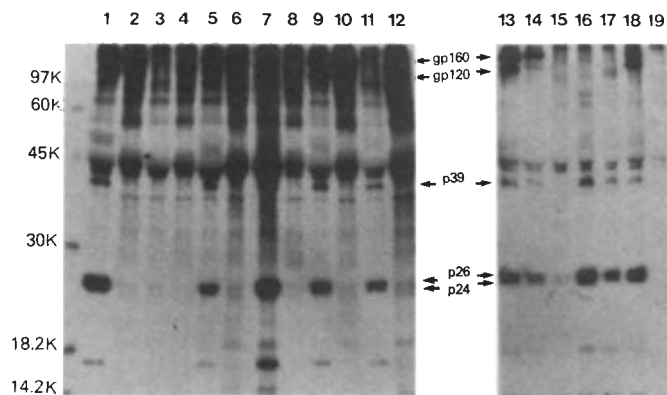


**Fig. 2** Propagation of JS4 viruses in T cells. COS-1 cells ( $5 \times 10^5$  to  $1 \times 10^6$ ) were transfected with plasmid DNA (5–10  $\mu$ g) and after 24 h combined with  $1 \times 10^6$  polybrene-treated ( $2 \mu$ g ml $^{-1}$  for 30 min) H9 cells (a), MOLT-3 cells (b), Jurkat cells (c) or cord blood mononuclear cells from normal donors (d). After 5 days non-adherent cells were removed and maintained independently in suspension culture in RPMI 1640 media containing 10% fetal calf serum. Cultures were examined for the expression of HIV-1 *gag* (p24) using the monoclonal antibody BT3<sup>19</sup>. The results shown in each of panels (a), (b), (c) and (d) summarize at least three independent experiments, for clones JS4-6 (■), JS4-15 (△), JS4-16 (○), JS4-22 (●), JS4-26 (□) and pHXB2gpt (▲) and duplicate experiments for clone JS4-24 (◇). Values between zero and <0.01% are shown as 0.01% to represent the lowest accurate level of detection.

**Methods.** Transfection and establishment of cocultures was carried out as described previously<sup>17,20</sup>. Indirect immunofluorescence assays were performed on acetone-methanol fixed cells (50%, 10 min). Cord blood mononuclear cells were stimulated with phytohaemagglutinin for 5 days and maintained in media supplemented with IL-2 (10%) thereafter.

hybrid viruses were similar, making such an interpretation unlikely.

To analyse HIV-1-specific protein expression by JS4 clones, we performed immunoprecipitation studies using lysates from COS-1 transfected cells. Expression of HIV-1 *gag* proteins by JS4 genomes was similar to that seen for the parental strain. As illustrated in Fig. 3, lysates from COS-1 cells transfected with JS4-6, 16, 22 and 26 (lanes 5, 7, 9, 11) showed prominent virus-specific bands corresponding to the HIV-1 *gag*-related products p24, p26 and p39, when precipitated with a high titre human sera obtained from an AIDS patient. Similar results were seen with clones JS4-15 and 24 (data not shown) confirming that the HIV-1 *gag* expression by the hybrid clones was comparable to that of the wild type (pHXB2gpt, lane 1). Expression of HIV-1 *env*-encoded products (gp160 precursor and gp120) appeared altered in three of the five JS4 clones analysed. JS4-6 and JS4-22 lysates showed prominent bands of relative molecular mass ( $M_r$ ) 120,000 (120K) and 160K (lanes 6 and 10) when precipitated with goat antisera raised to HIV-1 recombinant envelope, which were absent from control samples (lane 4). Comparable species could not be routinely visualized in lysates prepared from JS4-15, 16 and 26 although occasionally a faster migrating (<120K) species, which appeared to be HIV-1 specific, was detected using high titre patient sera (see lanes 16, 15 and 17 respectively). Failure to detect normal-sized *env*-derived products may result



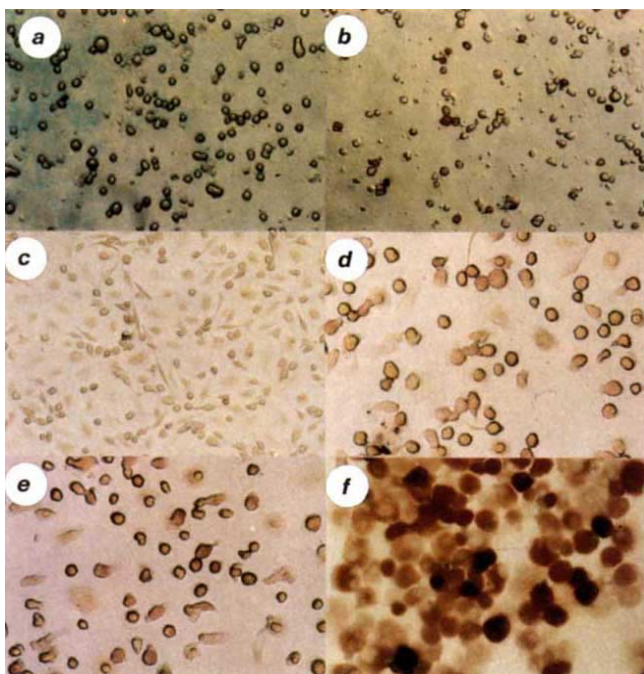
**Fig. 3** Radioimmunoprecipitation analysis of HIV-1 specific proteins produced by COS-1 cells transfected with JS4 genomes. Lysates of COS-1 cells following transfection with pHXB2gpt (lanes 1, 2 and 18), pSP65gpt (lanes 3, 4 and 19), JS4-6 (lanes 5, 6 and 14), JS4-15 (lane 16), JS4-16 (lanes 7, 8 and 15), JS4-22 (lanes 9, 10 and 13) and JS4-26 (lanes 11, 12 and 17) were precipitated with human sera obtained from an AIDS patient (lanes 1, 3, 5, 7, 9, 11 and 13–19), or with goat antisera raised against HIV-1 recombinant envelope protein (*env*9). Virus-specific *gag*- (p24, p26, p39) and *env*-derived products (gp120, gp160) are shown.

**Methods.** COS-1 cells ( $1 \times 10^6$ ) were transfected with 10  $\mu$ g of plasmid DNA and 48 h later labelled with [ $^{35}$ S]methionine and [ $^{35}$ S]cysteine (200  $\mu$ ci ml $^{-1}$ ) for 6 h. Cell lysates were prepared, processed and separated by SDS-PAGE (10%) as described elsewhere<sup>21</sup>.

from *env* truncation or modification, or from the envelope proteins having a reduced stability. At present, without the nucleotide sequence of these clones, it is not possible to distinguish between these possibilities but we note that clones which had previously failed to establish stable infection of T cells, (JS4-15, 16 and 26) were also those for which *env* products were unusual or could not be visualized.

The ability of JS4 viruses to replicate in monocytoid cells was investigated by inoculating human peripheral-blood derived monocyte-macrophage cultures (containing >99.99% peroxidase positive, non-specific esterase positive cells) with concentrated virus. Cultures were maintained in media supplemented with interleukin 3 (IL-3) and colony stimulating factor 1 (CSF-1) and stained for HIV-1 *gag* antigen expression 5 weeks after inoculation. As shown in Fig. 4, cultures receiving HXB2-derived virus failed to show evidence of HIV-1 *gag* expression in four separate experiments. This result is consistent with HXB2 being a component of the HTLV-IIIB isolate, one passaged extensively in T cells which appears to be very poorly infectious in monocytes<sup>10</sup>. In contrast, HIV-1 *gag*-expressing cells were detected in parallel cultures inoculated with virus derived from JS4-16 (Fig. 4d) and JS4-15 (Fig. 4e), and rare positive cells were observed in cultures derived from JS4-6 and JS4-22 (in three of four experiments). JS4-26 virus failed to establish infection under the conditions used and JS4-24 was not studied. Attempts to detect reverse transcriptase activity in the culture media removed at weekly intervals, showed low activity (at most  $3 \times$  background) which peaked at 5–6 weeks.

The finding that four out of the five JS4 viruses can be maintained in monocytoid cultures, whereas only one of the viruses (JS4-22) grows well in H9 cells, might suggest that the JS4 viruses show a tropism for monocytes. However, it could reflect the interplay between virus infection and cell division: if a virus such as JS4-15 were capable of replicating slowly in culture, infection would tend to be lost in rapidly dividing cultures (such as H9, where infected cells would be quickly diluted out), but would be more easily retained in cultures



**Fig. 4** Expression of HIV-1 by monocytoid cultures inoculated with JS4 viruses. Indirect labelling, using a monoclonal antibody to HIV-1 p24 (BT3)<sup>19</sup> and an alkaline phosphatase coupled anti-mouse Ig, was used to visualize HIV-1 infected cells in H9/HTLV-III<sub>B</sub> cultures (b), and monocytoid cultures which had been inoculated with JS4-16 (d) or JS4-15 (e) derived virus, five weeks previously. Parallel uninfected H9 cultures (a) and monocytoid cultures which had been inoculated with HXB2gpt derived virus (but failed to show evidence of HIV-1 expression) (c) are also shown. Uninfected cultures of monocytes (week 5) were stained for non-specific esterase (f) and peroxidase production to check that the populations were monocytoid.

**Methods.** Mononuclear blood cells were obtained from single normal donors and applied to a continuous percoll gradient to yield a cell population enriched for monocytes, as described elsewhere<sup>22</sup>. Cells were washed and plated at subconfluence ( $3-6 \times 10^5$  per well) in 24-well plates, and maintained in RPMI 1640 media containing 5% human AB serum, 10% fetal calf serum and antibiotics. After 6-24 h, nonadherent cells were removed by washing and fresh media containing lipopolysaccharide ( $10 \text{ ng ml}^{-1}$ ) was added to the cultures. To stimulate the monocyte/macrophages, recombinant CSF-1 (15 units per ml), IL-3 (5 units per ml) or conditioned media removed (day 16-21) from primary cultures of human cord blood endothelial cells (CM-ENDO) (10%) was also included in the culture medium. After 5 days the cultures were treated with polybrene ( $2 \mu\text{g ml}^{-1}$ , 30 min), rinsed with phosphate buffer saline (PBS) and  $100 \mu\text{l}$  of concentrated virus was applied to the cells for 1 h at  $37^\circ\text{C}$ . The cultures were fed with 2 ml fresh media (supplemented with IL-3+CSF-1 or 10% CM-ENDO) and maintained for 5 weeks (with weekly media changes) before being fixed *in situ* with 0.5% glutaraldehyde (in PBS) (5 min,  $20^\circ\text{C}$ ), rinsed, incubated with a glycine solution (100 mM with 0.1% bovine serum albumin, 30 min,  $20^\circ\text{C}$ ), rinsed and dried and stained. Virus used for inoculation was prepared by transfecting  $6 \times 10^7$  COS-1 cells with 300  $\mu\text{g}$  of plasmid DNA as described previously<sup>20</sup>, harvesting the culture supernatant after 7 days and concentrating virus particles by direct centrifugation ( $100,000g$  for 1 h at  $4^\circ\text{C}$ ). The amount of virus contained in each preparation was standardized by performing RT assays and direct particle counts (by electron microscopy). In a typical experiment the values for RT and particle counts for virus stocks HXB2, JS4-6, JS4-15, JS4-16, JS4-22 and JS4-26 were  $2.1 \times 10^{10}$  particles  $\text{ml}^{-1}/11,000$  d.p.m.;  $3.1 \times 10^{10}$  particles  $\text{ml}^{-1}/16,000$  d.p.m.;  $2.4 \times 10^{10}$  particles  $\text{ml}^{-1}/19,000$  d.p.m.;  $2.2 \times 10^{10}$  particles  $\text{ml}^{-1}/24,000$  d.p.m.;  $2.4 \times 10^{10}$  particles  $\text{ml}^{-1}/44,000$  d.p.m.; and  $2.3 \times 10^{10}$  particles  $\text{ml}^{-1}/49,000$  d.p.m., respectively.

showing only limited proliferation (such as CB cultures) and would be maintained in cultures which do not divide (such as monocytes). Alternatively, the failure of JS4-15, JS4-16 and JS4-26 to be efficiently transmitted in culture could reflect the rapidity with which they kill an infected cell. It is conceivable that poorly replicating or defective HIV-1 substrains (such as JS4-26) do exist *in vivo* and are extremely cytopathic, but require the assistance of other strains (such as JS4-22) to help in transmission. A recent report<sup>11</sup> in which replication-defective variants of feline leukaemia virus (FeLV) were found to initiate rapid fatal immunodeficiency in cats when co-inoculated with a non-pathogenic replication-competent helper virus<sup>11</sup>, adds credence to such a hypothesis. The finding that four of six clones analysed could not be maintained in Jurkat or CB cells (cells from which the viruses were originally cloned) is paradoxical. A possible explanation is that infected cells can accommodate more than a single viral prototype, and the co-existence of replication competent and defective clones with a single cell permits the transmission and survival of the defective form. Further experiments including the molecular analyses of DNA isolated from single cell clones will be needed to clarify this point.

In conclusion, we provide evidence that extensive biological variation exists *in vivo* in HIV-1 infected individuals and in isolates not subjected to extensive passage. It is worth contrasting the variation seen in the R.J.S. isolate with the close biological similarity of molecular clones derived from HTLV-III<sub>B</sub> (an isolate adapted to long term growth in H9 cells including clones HXB2, HX10, HXB3; ref. 12). The biological diversity, present *in vivo*, is presumably lost *in vitro* by the selective outgrowth of HIV-1 strains capable of rapid replication. This notion is supported by extensive studies on influenza virus (in which isolation of virus from fertilized hens' eggs and from canine kidney cultures, results in the generation of two antigenically distinct populations)<sup>13,14</sup> and by reports of host restriction in visna virus<sup>15</sup>. Our data also suggest that monocytoid cultures may be extremely useful for the future identification and propagation of slow-replicating HIV-1 strains. In addition, it will be of considerable importance to clarify how accurately *in vitro* adapted HIV-1 strains reflect the properties of *in vivo* strains and whether defective or slow replicating viruses have significance in disease progression.

We thank Drs Robert Redfield and Philip Markham for help in obtaining samples for study, Drs Matthew Gonda, Martin White and John Lemp for help with electron microscopy, Ms Randy Zicht for preparation of monocytes and Dr Ann Bodner, Lynn Simek, Sharon Kensek and Piera Cassettari for additional assistance. We are extremely grateful to Dr Steven Clark for the recombinant CSF-1 and to Drs L. Ivanoff and S. Petteway for high titre goat antisera to recombinant HIV-1 envelope.

Received 1 March; accepted 29 June 1988.

- Gottlieb, M. S. *et al.* *New Engl. J. Med.* **305**, 1424-1431 (1981).
- Gallo, R. C. *et al.* *Science* **224**, 500-502 (1984).
- Barre-Sinoussi, F. *et al.* *Science* **220**, 868-871 (1983).
- Clavel, F. *et al.* *Science* **233**, 343-346 (1986).
- Shaw, G. M. *et al.* *Science* **226**, 1165-1171 (1984).
- Hahn, B. H. *et al.* *Nature* **330**, 184-186 (1987).
- Hahn, B. *et al.* *Science* **232**, 1548-1553 (1986).
- Saag, M. *et al.* *Nature* (this issue).
- Fisher, A. G. *et al.* *Nature* **320**, 367-371 (1986).
- Gartner, S., Markovitz, P., Markovitz, D. M., Betts, R. F. & Popovic, M. *Science* **233**, 215-219 (1986).
- Overbaugh, J., Donahue, P. R., Quackenbush, S. L., Hover, E. A. & Mullins, J. I. *Science* **239**, 906-910 (1988).
- Ratner, L. *et al.* *Aids Res. hum. Retroviruses* **3**, 57-69 (1987).
- Robertson, J. S. *et al.* *Virology* **160**, 31-37 (1987).
- Katz, J. M., Naeve, C. W. & Webster, R. G. *Virology* **156**, 386-395 (1987).
- Haase, A. T., Stowring, L., Narayan, O., Griffin, D. & Price, D. *Science* **195**, 175-177 (1977).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456-467 (1973).
- Rho, H. M., Poiez, B., Ruscetti, F. & Gallo, R. C. *Virology* **112**, 355-360 (1981).
- Vernese di Marzo, F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 5199-5205 (1985).
- Fisher, A. G. *et al.* *Science* **237**, 888-893 (1987).
- Barin, F. *et al.* *Science* **228**, 1094-1098 (1985).
- Kleinerman, E. S. *et al.* *Cancer Res.* **43**, 2010-2014 (1983).



# The helical repeat of double-stranded DNA varies as a function of catenation and supercoiling

Steven A. Wasserman\*†, James H. White‡  
& Nicholas R. Cozzarelli\*

\* Department of Molecular Biology, University of California, Berkeley, California 94720, USA

‡ Department of Mathematics, University of California, Los Angeles, California 90024, USA

DNA in the cell is intertwined at several levels: one polynucleotide strand wraps helically around its complement and the double helix is in turn coiled in space. The higher-order intertwining most often takes the form of supercoiling of the helix axis<sup>1</sup>, but can also be observed as the wrapping of one DNA duplex around another, as in catenation<sup>2,3</sup>. We have investigated the relationship between intertwining at these three levels, the double helix, supercoiling, and catenation, using an approach that relies on comparative measurements of DNA linking numbers by gel electrophoresis. The method determines both the handedness of DNA catenanes and the change in helical repeat that accompanies catenation-induced supercoiling. For multiply-linked catenated rings of 3.5 kilobase pairs (kb), we conclude that the double helix unwinds by two-thirds of a turn for every right-handed supercoil involved in linking the two circles. Altering the geometry of the catenanes by linking rings of dissimilar size changes the effect of catenation on helical and superhelical parameters. Our experiments used intact DNA rings, but we note that linear DNA molecules, by virtue of their subdivision into closed loops or domains *in vivo*, can intertwine in the same ways<sup>4</sup>.

The linking number of naturally occurring DNA is less than that of DNA that has been relaxed by nicking and religation or by treatment with a topoisomerase. The linking difference,  $\Delta Lk$ , is therefore negative. Although such DNA is said to be negatively supercoiled, it is clear that  $\Delta Lk$  can also take the form of an altered helical winding because the linking number, as considered in this report, is the sum of the number of times either polynucleotide strand is wound about the helix axis and the number of times the helix is wound around the supercoil axis<sup>5,6</sup>. Therefore, in investigating the structural consequences of intertwining we have analysed both supercoiling and helical winding.

We begin by considering intertwining in the form of a winding of one DNA duplex around another. In the example shown in Fig. 1a, the two catenated duplexes are linked by left-handed intertwinings and are supercoiled in a left-handed (negative) fashion. Similarly, the right-handed intertwinings in Fig. 1b result in a positive supercoiling of each ring. These catenanes and the supercoils are said to be toroidal because they can be constructed by a non-intersecting winding around a torus, or doughnut-shaped object.

If the component rings of a catenane such as that in Fig. 1 are nicked and then resealed, only that supercoiling inherent to catenation should remain. If one ring of the catenane is then linearized, the handedness of the catenane might be evident from the sign of the linking difference for the intact ring with respect to an identical ring that was not catenated when religated. This approach is experimentally analogous to the determination of the handedness of the DNA wrapped in nucleosomes, where the measured linking difference reflected in part the left-hand coiling of DNA around histones<sup>7</sup>.

The linking difference measured in such an experiment is a function of both the number of catenane intertwinings and the

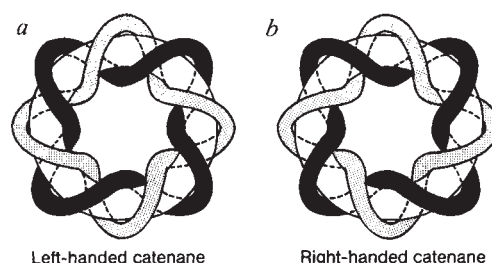


Fig. 1 Torus catenanes. The black and grey ribbons represent circular duplex DNA molecules that are wrapped around each other helically and lie without intersection on the surface of a torus. In the catenanes shown, the duplexes wrap around each other and the torus axis four times. This quantity, the linking number of the torus axis and a DNA axis, is the supercoiling number for each ring<sup>5</sup>. In (a), each duplex has four negative toroidal supercoils, whereas the four supercoils in (b) are positive. Topological sign is determined by the usual convention<sup>3</sup>.

helical repeat of duplex DNA. For every intertwining added, another supercoil is required in each catenated ring (see Fig. 1). The change in linking inherent in this supercoiling will be modulated by any change in helical repeat that results from the bending of the helix axis into a superhelical form. Indeed, according to the linking number relationship above, the difference between  $\Delta Lk$  and the supercoiling winding number, as deduced here from the number of catenane intertwinings, equals the change in helical winding upon intertwining.

Dimeric torus catenanes, the class of catenanes commonly found *in vivo*<sup>2,3</sup>, can be readily generated by site-specific recombination *in vitro*. The Int proteins (lambda integrase and *Escherichia coli* integration host factor) mediates recombination of supercoiled DNA substrates containing bacterial and lambda phage attachment sites (*attB* and *attP*)<sup>8</sup> to yield multiply intertwined catenanes that are exclusively right-handed<sup>9</sup>.

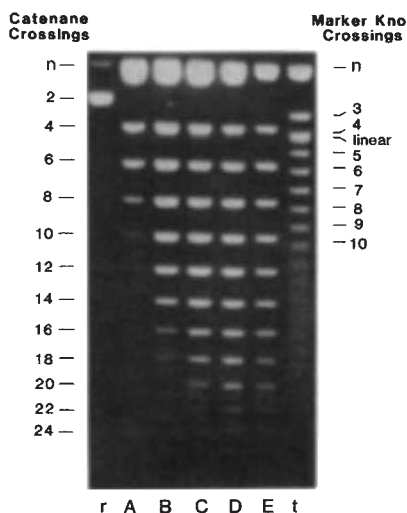
A second recombination system, that of Tn3 resolvase, generates singly-intertwined catenanes<sup>10</sup>. The recombination substrate pAB7.0d can be reacted with either Int or resolvase to yield torus catenanes composed of two rings of nearly equal size, each with unique restriction sites.

To examine the relationship between catenation and net distortion of the double helix, we prepared Int catenanes that varied in their degree of intertwining by varying the superhelical density of the substrate<sup>3</sup>. Samples of pAB7.0d of increasing superhelical density were generated using a type I topoisomerase to nick and reclose DNAs in the presence of increasing concentrations of ethidium bromide. After reacting these samples with Int, each ring was nicked once and the catenanes were analysed by agarose gel electrophoresis. As shown in Fig. 2, the mean number of product intertwinings increased in proportion to the  $\Delta Lk$  of the Int substrates from a low linking deficit (A) through one slightly greater than found *in vivo* (D).

Using the nicked catenated DNAs generated with Int and DNaseI, we prepared two sets of samples for electrophoretic analysis. To make precut samples, one ring was linearized before ligation of the nicks; to make postcut samples, linearization took place after ligation. In the presence of chloroquine to positively supercoil the DNA, the resealed ring from the postcut catenanes migrated, on average, faster than that from the precut ones for all preparations (Fig. 3). Thus the right-handed intertwining of the catenated DNA products was in fact reflected in an increased linking number.

In Fig. 4a, we graph  $\Delta Lk$  against the average number of catenane intertwinings for pAB7.0d, using the data from several experiments. The plot is linear over the range of values measured, indicating a constant change in DNA structure for each additional intertwining. The slope of the line in Fig. 4a shows that each intertwining or supercoil increased  $Lk$  by only

† Present address: Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA.



**Fig. 2** Electrophoretic analysis of catenane intertwining. Lanes A–E: 300 ng of nicked, multiply-intertwined pAB7.0d catenanes generated with Int. Lane r: 80 ng nicked singly-linked pAB7.0d catenanes generated with resolvase. Lane t: 750 ng electrophoretic standards (topoisomerase-generated pAB7.0d knots).

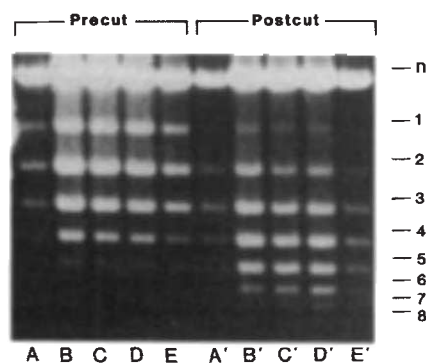
**Methods.** DNA samples were electrophoresed for 50 h at  $1.5 \text{ V cm}^{-1}$  in an 0.9% agarose gel in SDS-containing Tris-acetate buffer<sup>2</sup>. Nicked DNAs of identical size migrate according to the number of duplex intertwinings, and knots migrate slightly faster than catenanes with the same number of crossings<sup>3</sup>. The catenated Int products show a steps-of-two spacing, whereas the spacing of the reference knots is steps-of-one. The linear species is a by-product of *DNase I* nicking; nicked starting material is denoted by n. Int substrates differing in *Lk* were generated by treating pAB7.0d with wheat germ topoisomerase I in the presence of 0.75 (A), 1.5 (B), 2.25 (C), 3.0 (D), or 3.75 (E)  $\mu\text{g ml}^{-1}$  ethidium bromide. The DNAs were then treated with Int<sup>15</sup> and nicked with *DNase I*. Singly-linked catenanes were generated by reacting supercoiled pAB7.0d with Tn3 resolvase<sup>10</sup> and then nicking the products with *DNase I*. Marker knots were prepared with phage T4 topoisomerase<sup>17</sup> and then nicked with *DNase I*.

about one-third of a turn. Thus, the helix must unwind by approximately two-thirds of a turn for each supercoil added. Assuming an initial value of 10.5 base pairs (bp) per turn, the helical repeat increases by 0.1 bp per turn for every five intertwines.

Unlike subsequent intertwinings, the first one between two DNA circles need not distort the DNA, as singly-linked rings can still adopt a planar conformation. Indeed, the extrapolated linear regression plot in Fig. 4a suggested that a preparation of such catenanes would show no linking deficit. We tested this directly, analysing the conformation of the singly-linked 3.0 and 4.0 kb rings generated by reacting pAB7.0d with the Tn3 resolvase (Fig. 3, lane r). As expected, no net  $\Delta Lk$  was detectable in either ring (data not shown).

The *Lk* differences detected with multiply-linked duplexes indicate a mutual distortion of the catenated rings. We investigated the relationship between catenane intertwining,  $\Delta Lk$ , and helical repeat in a situation where the overall geometry is very different. A 2.9-kb ring, similar in size to the pAB7.0d circles, was catenated to a ring of only 900 bp. Because of the substantial energetic barrier to supercoiling of DNA rings less than one kb in size<sup>11,12</sup>, the smaller ring should be less readily distortable. These catenanes were produced by reacting Int with a second substrate, plasmid pBNW3.8d, and should consist of a 2.9-kb DNA circle winding around a roughly planar 0.9 kb ring.

pBNW3.8d DNAs varying in *Lk* were prepared and reacted with Int. The catenanes were then nicked and the linking difference entrapped in each ring determined. As with pAB7.0d, the right-handed catenane intertwinings were reflected in a posi-



**Fig. 3** Electrophoretic analysis of DNA supercoiling induced by catenation.

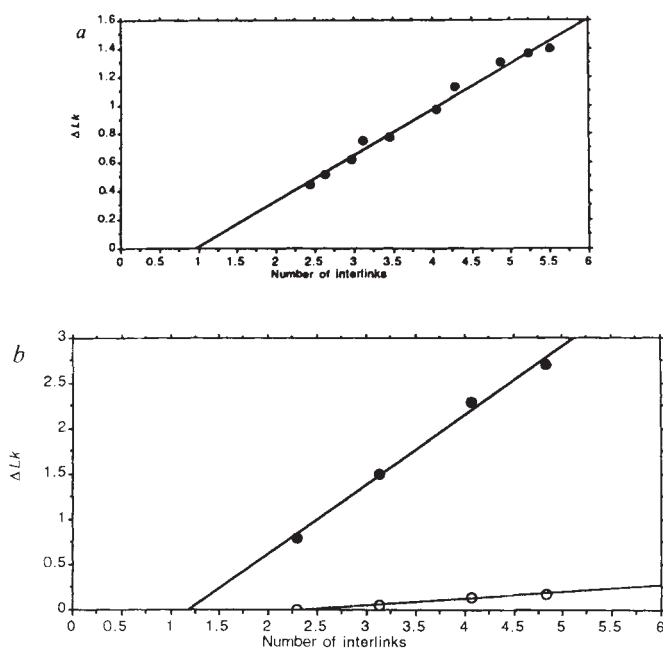
**Methods.** Samples (2.5  $\mu\text{g}$ ) of DNA mixtures A–E (Fig. 2) were incubated in 14  $\mu\text{l}$  buffer P (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM  $\text{MgCl}_2$ , 10 mM dithiothreitol, 1 mM ATP, and 50  $\mu\text{g ml}^{-1}$  bovine serum albumin) with *Bam*HI (precut samples) or *Bam*HI storage buffer alone (postcut samples). After 30 min at 32 °C, 2  $\mu\text{l}$  of T4 DNA ligase in buffer P was added and the samples were incubated for an additional 30 min. The ligase was inactivated by heating for 10 min at 65 °C, after which the samples were cooled to 32 °C and incubated for 30 min with 2  $\mu\text{l}$  of buffer P containing either *Bam*HI (postcut samples) or *Bam*HI storage buffer (precut samples). Electrophoresis of precut (A–E) and postcut (A'–E') samples was carried out for 24 h at  $2 \text{ V cm}^{-1}$  in an 0.8% agarose gel in Tris-acetate buffer containing 1  $\mu\text{g ml}^{-1}$  chloroquine. That portion of the ethidium bromide-stained gel containing topoisomers of the uncut recombinant ring is shown. The nicked monomer is denoted as n; values at right indicate topoisomer number; the topoisomers shown are positively supercoiled. The shape of the postcut distributions reflects non-Gaussian spread in intertwining number for the Int catenanes.

tive  $\Delta Lk$  whose magnitude was linearly proportional to the number of catenane intertwinings (Fig. 4b). Per base pair, there was a greater than three-fold preference for distributing  $\Delta Lk$  to the 2.8-kb ring, indicating that supercoiling is for the most part restricted to the large ring. The slope of the  $\Delta Lk$  plot for the 2.8-kb circle (Fig. 4b) shows that the double helix comprising this ring unwinds about one-quarter of a turn per catenane intertwining. This small degree of unwinding is less than half that seen with pAB7.0d catenanes possessing a similar number of intertwinings, presumably because the 2.8-kb circle has fewer points of contact and constraint than it would if catenated to a ring whose circumference were closer to its own.

Our experiments show that: (1) catenation of DNA duplexes results in a  $\Delta Lk$  after nicking and religation whose magnitude varies linearly with the number of intertwinings and whose sign is predicted by the handedness of the intertwining; (2) this  $\Delta Lk$  is a result of both writhing and an altered helical repeat—the right-handed toroidal supercoiling resulting from catenation unwinds the double helix in direct proportion to the number of supercoils; and (3) singly-intertwined catenanes do not result in a change of *Lk* and therefore only multiple-intertwining provides supercoiling energy for processes influenced by DNA structure.

Previously it had only been possible to determine catenane handedness by electron microscopy of DNA coated with the *recA* protein<sup>9</sup>. The electrophoretic method described here depends on the overall catenane population rather than individual molecules, provides insight into catenane structure in solution, and can be used with catenanes too complex for ready viewing. Nevertheless, more complete characterization requires electron microscopy of molecules for which *RecA* coating and partial denaturation have made visible both the relative orientation of the intertwined rings and the handedness of intertwining<sup>3</sup>.





**Fig. 4** Dependence of entrapped linking difference on catenane complexity. The abscissa indicates the average number of intertwinings for the catenated fraction of each product mixture. The ordinate plots the average shift in  $Lk$  observed for the uncut DNA ring. Data gathered as described in Figs 2 and 3 were quantitated with a Hoeffer Model GS3000 densitometer.  $\Delta Lk$  was calculated by subtracting the mean topoisomer number for each precut sample from the mean for the corresponding postcut DNA. *a*, pAB7.0d catenanes. The slope of the least squares regression line ( $r^2 = 0.99$ ) is 0.32; the x-intercept is 0.95. *b*, pBNW3.8d catenanes. (●)  $\Delta Lk$  for 2.9 kb ring; the slope of the regression line is 0.76. (○)  $\Delta Lk$  for 0.9 kb ring; the regression line has a slope of 0.073.

That supercoiling *per se* can alter the helical repeat of duplex DNA independent of protein-DNA interactions was previously suggested on theoretical grounds by Levitt<sup>13</sup>. He argued that a decrease in the helical repeat would be needed to accommodate an increased base separation at the compacted inner side of a duplex bend. A decrease in helical repeat was subsequently confirmed for nucleosome-bound DNA<sup>14</sup>, but does not hold true for the catenanes studied here in which the repeat actually increases. Although the supercoils modelled by Levitt had a handedness opposite to that of the right-handed catenanes, the need to relieve stress on the inner side of the bends in DNA should be about the same for both structures. If, however, the change in helical repeat with supercoiling results from the geometry of DNA winding, rather than steric interactions between bases, the observed opposite changes in helical repeat for right-handed catenanes and left-handed nucleosomal DNA are predicted<sup>6</sup>.

The method we devised to measure helical repeat depends on the number of supercoils of each catenated ring being equal to the readily measurable number of intertwinings between rings. The results apply only to toroidal supercoils as also occur in nucleosomal DNA, but not to the plectonemic (interwound) forms that predominate for underwound DNA in solution<sup>15</sup>. But we expect on theoretical grounds that plectonemic supercoiling will also change the repeat of the double helix<sup>6</sup> and that there is thus a general relationship between helical parameters and DNA supercoiling. Bending of DNA can change helical repeat just as changes in the structure of the double helix can cause bending.

Supercoils, catenanes and knots, the three common higher-order forms of DNA intertwining<sup>3</sup>, can have similar bends and

crossings, despite differing in global connectivity. The results presented here support the idea that these forms share a common local structure in terms of altered helical repeat. That catenane or knot intertwinings can substitute for plectonemic supercoiling in directing DNA transposition has been demonstrated directly by Craigie and Mizuuchi<sup>16</sup>. Moreover, multiple (H. Benjamin, and N.R.C., unpublished data) but not single<sup>10</sup> intertwining of catenanes directs site specific recombination by Tn3 resolvase, and we find an induced  $\Delta Lk$  in DNA only for multiply-intertwined catenanes.

The distortion of helical geometry in multiply-intertwined DNAs could provide a directional energetic drive for many DNA transactions *in vivo* such as the activity of transcription and replication factors. The magnitude of unwinding caused by intertwining can be quite large. For pAB7.0d, helical repeat increased by 0.1 for every five supercoils. The analogous right-handed wrapping of 100 bp of DNA around a protein to give one positive supercoil would shift the helical repeat from 10.5 to 11.3. An increase in repeat of similar proportion in the transition from B-form to A-form duplex is accompanied by a 20° base-pair tilt. The results here therefore suggest that intertwining, whether in the form of catenation or supercoiling alone, profoundly alters the architecture, and hence the activity, of the DNA double helix.

J. Bliska and A. Hemmati-Brivanlou shared in the plasmid construction. We also thank A. Varshavsky for his catalytic role, K. Kreuzer for topoisomerase, M. Krasnow for communication of preliminary results, C. Boles and F. Dean for helpful discussion, and J. Jenkins for preparation of the manuscript. S.A.W. is a Lucille P. Markey Scholar. This work was supported by a grant from the Lucille P. Markey Charitable Trust and by NIH and NSF grants to N.R.C. and J.H.W.

Received February 18; accepted 26 June 1988.

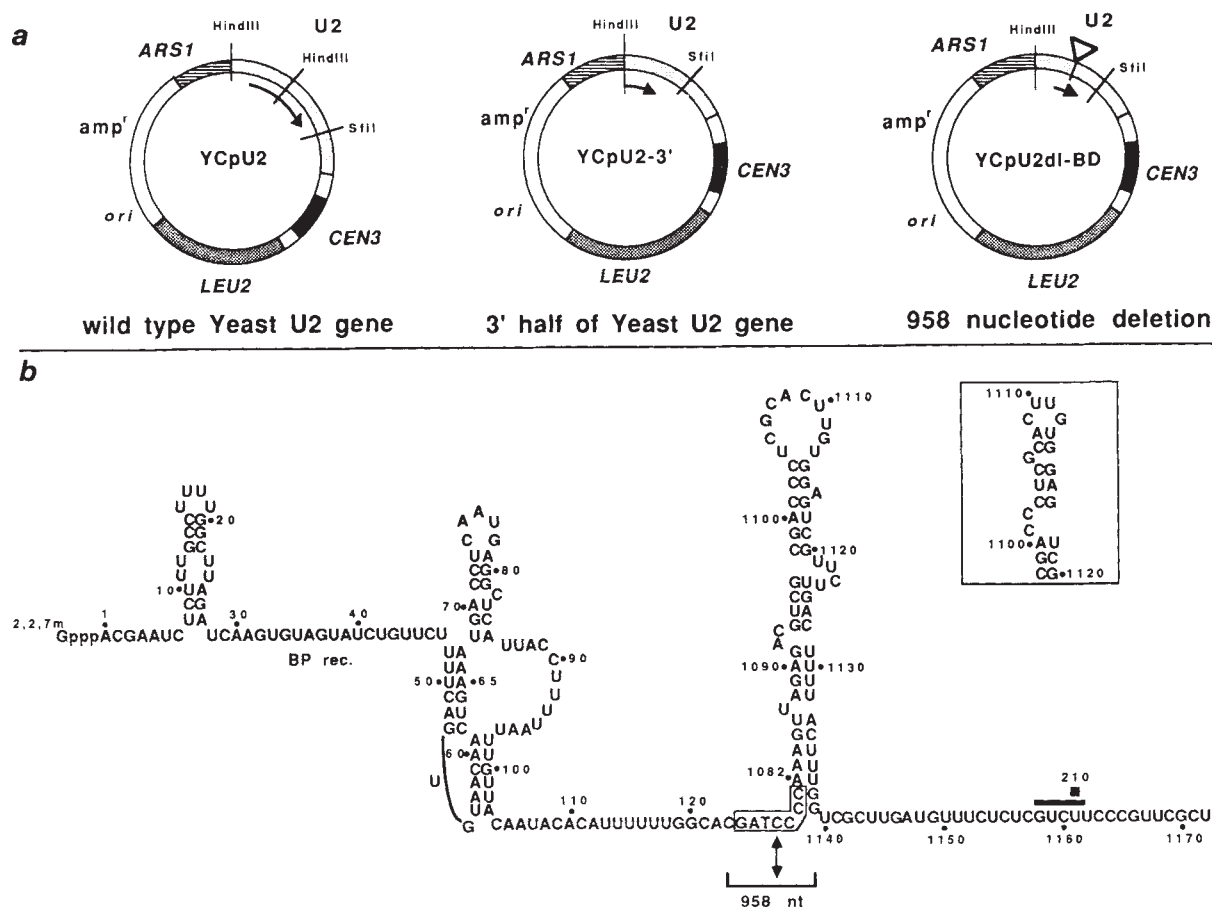
1. Wang, J. C. A. *Rev. Biochem.* **54**, 665-697 (1985).
2. Sundin, O. & Varshavsky, A. *Cell* **21**, 103-114 (1980); **25**, 659-669 (1981).
3. Wasserman, S. A. & Cozzarelli, N. R. *Science* **232**, 951-960 (1986).
4. Worcel, A. & Burgi, E. *J. molec. Biol.* **71**, 127-147 (1972).
5. Vinograd, J., Lebowitz, J. & Watson, R. *J. molec. Biol.* **33**, 173-192 (1968).
6. White, J. H., Cozzarelli, N. R. & Bauer, W. R. *Science* **241**, 323-327 (1988).
7. Germond, J. E., Hirt, B., Oudet, P., Gross-Bellard, M. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1843-1847 (1975).
8. Nash, H. A. A. *Rev. Genet.* **15**, 143-167 (1981).
9. Spengler, S. J., Stasiak, A. & Cozzarelli, N. R. *Cell* **42**, 325-334 (1985).
10. Krasnow, M. A. & Cozzarelli, N. R. *Cell* **32**, 1313-1324 (1983).
11. Horowitz, D. S. & Wang, J. C. *J. molec. Biol.* **173**, 75-91 (1984).
12. Shore, D. & Baldwin, R. L. *J. molec. Biol.* **170**, 983-1007 (1983).
13. Levitt, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 640-644 (1978).
14. Klug, A. & Lutter, L. C. *Nucleic Acids Res.* **9**, 4266-4283 (1981).
15. Bliska, J. B. & Cozzarelli, N. R. *J. molec. Biol.* **194**, 205-218 (1987).
16. Craigie, R. & Mizuuchi, K. *Cell* **45**, 793-800 (1986).
17. Kreuzer, K. N. & Jongeneel, C. V. *Meth. Enzy.* **100**, 144-160 (1983).

## Internal sequences that distinguish yeast from metazoan U2 snRNA are unnecessary for pre-mRNA splicing

A. Haller Igel & Manuel Ares Jr

Biology Department, Thimann Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064, USA

U2 small nuclear RNA is a highly conserved component of the eukaryotic cell nucleus involved in splicing messenger RNA precursors<sup>1-3</sup>. In the yeast *Saccharomyces cerevisiae*, U2 RNA interacts with the intron by RNA-RNA pairing between the conserved branchpoint sequence UACUAAC and conserved nucleotides near the 5' end of U2 (ref. 4). Metazoan U2 RNA is less than 200 nucleotides in length<sup>5</sup>, but yeast U2 RNA is 1,175 nucleotides long<sup>6</sup>. The 5' 110 nucleotides of yeast U2 are homologous to the



**Fig. 1** Structure of plasmids and the U2-bd transcript. *a*, Plasmid constructs used in this study. Plasmids were derived by standard techniques<sup>21</sup> from plasmid R489 (gift of S. Roeder, Yale) which contains a *Sau3A* fragment of *CEN3* cloned in the *Bam*HI site, a *Hpa*I-*Sal*I fragment of the *LEU2* gene cloned between the *Sal*I and *Pvu*II sites, and an *Eco*RI-*Hind*III fragment of *ARS1* cloned between the (filled-in) *Eco*RI and *Hind*III sites of pBR322. A 1.3-kb *Hind*III fragment containing the 3' half of the yeast U2 gene (known as *LSR1* (ref. 6) or *SNR20* (refs 4, 22)) was introduced into the *Hind*III site and then the downstream *Hind*III site was destroyed by filling in with Klenow, producing YCpU2-3' (centre). YCpU2 was constructed by cloning the 0.95-kb *Hind*III fragment containing the promoter and 5' half of the yeast U2 gene into the *Hind*III site of YCpU2-3' in the correct orientation. YCpU2dl-BD was generated beginning with a subclone containing the *Dra*I-*Sma*I fragment in the *Sma*I site of mp10 (ref. 6). The *Bam*HI site adjacent to the fused *Dra*I-*Sma*I site was cleaved, filled in and the DNA was recut further down in the polylinker at the *Hind*III site. Into this site was ligated a U2 fragment cut at the *Ban*I site at position 119, filled in and recut upstream of the promoter with *Hind*III. This generated a subclone containing the promoter and coding sequences to position 123 fused to 7 bp derived from the polylinker (boxed sequence in Fig. 1*b*) fused to position 1,082 of the coding region (the *Dra*I site) and 3' flanking sequences. The sequence was confirmed across the junction by dideoxy sequencing<sup>23</sup>. This gene was introduced into YCpU2-3' using the unique *Hind*III and *Sfi*I sites. Arrows represent the U2 coding region. *b*, Secondary structure model of transcript coded by the U2dl-BD gene. Numbering begins with the first nucleotide. Pairings in the first 100 nt, including the pseudoknot, are phylogenetically conserved<sup>14</sup>. The 3' sequences can be folded to resemble either of two structures found in other U2 RNAs, stem IV (shown) or stem III (inset). Nucleotides derived from the polylinker are boxed in the sequence. Bracket indicates the site of the 958 nucleotide deletion. Bar above the sequence indicates region of 3' end formation of transcripts from the deletion gene. Square indicates position 210 of the U2-bd transcript (wild-type transcripts would extend past position 1,170 as shown).

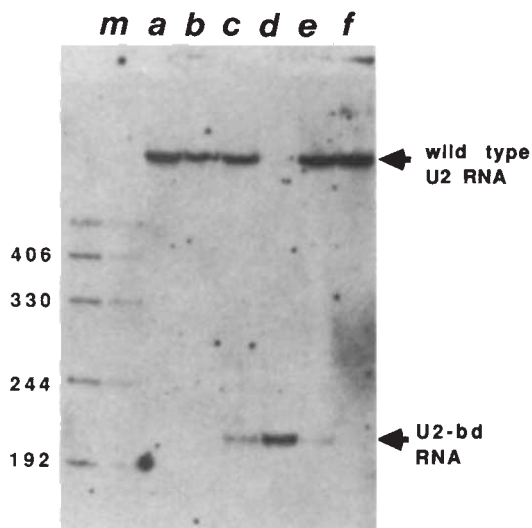
5' 100 nucleotides of metazoan U2 (ref. 6), and the very 3' end of yeast U2 bears a weak structural resemblance to features near the 3' end of metazoan U2. Internal sequences of yeast U2 share primary sequence homology with metazoan U4, U5 and U6 small nuclear RNA (ref. 6), and have regions of complementarity with yeast U1 (ref. 7). We have investigated the importance of the internal U2 sequences by their deletion. Yeast cells carrying a U2 allele lacking 958 nucleotides of internal U2 sequence produce a U2 small nuclear RNA similar in size to that found in other organisms. Cells carrying only the U2 deletion grow normally, have normal levels of spliced mRNA and do not accumulate unspliced precursor mRNA. We conclude that the internal sequences of yeast U2 carry no essential function. The extra RNA may have a non-essential function in efficient ribonucleoprotein assembly or RNA stability. Variation in amount of RNA in homologous structural RNAs has precedence in ribosomal RNA<sup>8,9</sup> and RNaseP<sup>10</sup>.

Although nuclear precursor mRNA (pre-mRNA) splicing in

*S. cerevisiae* and metazoans share many features, the yeast spliceosome appears to contain much more RNA than the metazoan spliceosome<sup>7,11</sup>, due primarily to the greater length of yeast U1 and U2 RNA. In yeast, U1 is 568 nucleotides (nt)<sup>7,11</sup> and U2 is 1,175 nt<sup>6</sup> long, whereas in metazoans U1 is 165 nt and U2 is 189 nt<sup>5</sup>. The discovery of the yeast homologues of U5 (ref. 12) and U4 plus U6 small nuclear RNAs (snRNA)<sup>13</sup> has made the notion obsolete that the extra RNA found in yeast U2 is arranged as a 'poly-snRNA' (ref. 6), but the function of the extra RNA remains obscure.

Conservation of U2 sequences in different organisms allowed the alignment of the 5' 120 nt of yeast U2 with the 5' 110 nt of metazoan U2 (ref. 6), but the relation of the remaining 1,050 nt of yeast U2 to the remaining 78 nt of metazoan U2 was not clear. Our analysis of the 3' end of yeast U2 using phylogenetic comparisons<sup>14</sup> or computer programs (also D. Konings; E. Shuster and C. Guthrie, personal communication) showed that the very 3' end of yeast U2 could fold into either of two potential



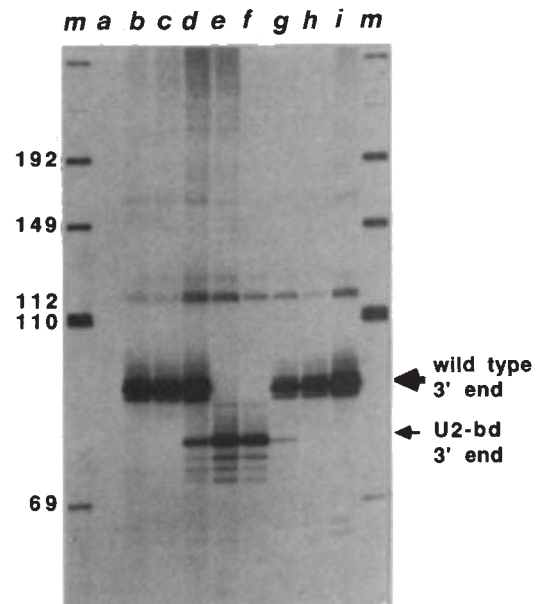


**Fig. 2** Size of U2 RNAs in different yeast strains. RNA was extracted, separated on a 0.4 mm thick 6% polyacrylamide-7M urea gel, transferred to nylon, and probed as described previously with kinase-labelled L15 oligonucleotide<sup>6</sup>. Samples are *a*, wild type, diploid MA3: *HIS4/his4-619 trp1/TRP1, leu2-3,112/leu2-3,112,ura3-52/ura3-52*; *b*, MA31: MA3 containing the U2::*URA3* disruption allele; *c*, MA31/YCpU2dl-BD: MA31 containing the YCpU2dl-BD plasmid; *d*, 503d: meiotic segregant of MA31/YCpU2dl-BD containing the U2::*URA3* disruption in the chromosome and the YCpU2dl-BD plasmid; *e*, 503a: spore (from same tetrad as 503d) containing the wild-type U2 allele in the chromosome and YCpU2dl-BD plasmid; *f*, 503a-LP: mitotic segregant of 503a that has lost the plasmid; *m*, markers (two lanes), *Hpa*II fragments of pUC13 labelled by filling in with Klenow. The genetic constitutions described for each strain used in these studies was confirmed by hybridization of blots of genomic DNA from each strain with a U2 probe.

structures, each similar to those found near the 3' end of vertebrate U2 (see Fig. 1). This suggests that the extra RNA could interrupt an otherwise typical U2 snRNA.

To test the functional importance of the internal sequences, we constructed a deletion to remove them. This deleted U2 gene (U2dl-BD) lacks 958 base pairs (bp) between positions 123 and 1,082 of the wild-type U2 gene and would be expected to encode a 224-nucleotide U2 molecule containing the highly conserved 5' end of U2 and the less well conserved 3' end (Fig. 1*b*). Three U2 gene constructs were incorporated into a yeast centromere plasmid (Fig. 1*a*): the complete wild-type yeast U2 gene (YCpU2), a U2 gene lacking the promoter and 5' half of the coding region (YCpU2-3'), and the U2dl-BD allele (YCpU2dl-BD).

We tested the ability of the three plasmids to rescue cells containing a lethal U2 mutation by introducing them into a diploid strain carrying one wild-type U2 gene, and one disrupted by deletion and insertion of the *URA3* gene<sup>6</sup>. We sporulated the diploids containing the plasmids and dissected tetrads. During meiosis, a centromere plasmid behaves in many respects like a normal chromosome<sup>15</sup>. If the plasmid-borne U2 gene is not functional, two of the meiotic products (spores) receive the chromosomal wild-type allele and survive (and are *ura*<sup>+</sup>), whereas the two that receive the disrupted U2 allele<sup>6</sup> die regardless of whether or not they receive the plasmid. If the plasmid-borne U2 gene is functional, spores that receive the chromosome with the lethal U2 allele have a chance to survive if they also receive the plasmid at meiosis. The results (not shown) reveal that the plasmid lacking the U2 promoter and 5' half of the U2 gene did not complement the disruption (no tetrads with >2 spores, no *ura*<sup>+</sup> spores), but both the wild-type U2 gene and the internal deletion did complement the disruption (tetrads

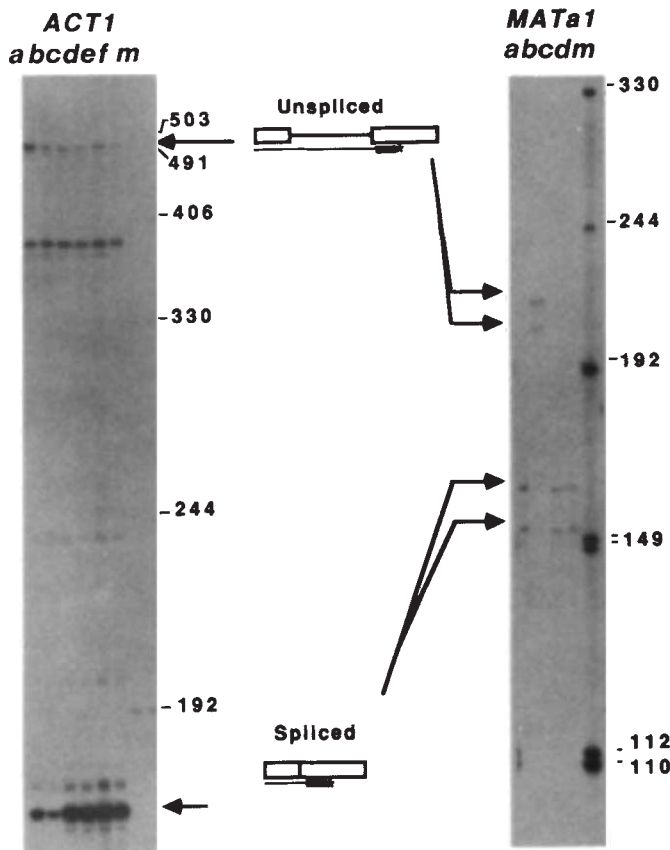


**Fig. 3** Mapping 3' ends of U2 transcripts in different yeast strains. An anti-sense probe transcript spanning from downstream of the gene and extended to position 1,082 in the coding region was synthesized using <sup>32</sup>P-labelled rGTP, and purified as described previously<sup>6</sup>. Two  $\mu$ g total yeast RNA was annealed to  $5 \times 10^4$  c.p.m. of probe and was digested with  $2 \mu\text{g ml}^{-1}$  RNase T1 at 30 °C, extracted and run out on a gel as before<sup>6</sup>. Samples are: (see Fig. 2 legend) *a*, digestion of probe after mock annealing with no RNA; after annealing to RNA from strains *b*, MA3; *c*, MA31; *d*, MA31/YCpU2dl-BD; *e*, 503d; *f*, H12: a derivative of 503d that has replaced the U2::*URA3* disruption allele with the U2dl-BD allele (a mitotic gene conversion event using the plasmid as a donor, obtained by selection against *URA3* on 5-fluoro orotic acid<sup>24</sup>) and then lost the plasmid; *g*, 503a; *h*, 503a-LP; *i*, 50a: a spore containing the U2::*URA3* allele on the chromosome and the wild-type U2 gene plasmid YCpU2; *m*, markers are *Hpa*II fragments of pUC13. The presence of the 120-nt fragment is variable and its origin has not been determined.

with >2 spores, some *ura*<sup>+</sup>). This genetic test shows that 958 nt of yeast U2 are not essential for function. Cells carrying only the deletion U2 gene grew normally. (Similar results on the dispensability of internal segments of the yeast U2 gene have been obtained by E. Shuster and C. Guthrie, personal communication.)

On the basis of the position of the 5' and 3' ends of wild-type U2 (ref. 6), we expected the U2dl-BD gene to encode a 220–224-nt transcript. Figure 2 shows that YCpU2dl-BD directs the accumulation of a transcript (U2-bd) that is close to this size, but significantly shorter at ~210 nt (Fig. 2). The 1,175-nt wild-type U2 transcript is absent from cells carrying U2dl-BD as the sole functional U2 gene, and the 210-nt transcript is the only detectable U2 RNA (Fig. 2, lane *d*), thus confirming the genetic conclusion that the central 958 nt are dispensable.

To determine the basis for the size difference, we mapped the 3' ends of U2 RNA contained in different strains using an RNase protection assay (Fig. 3). As expected, RNA from cells containing the wild-type U2 gene protects an 89-nt fragment of the probe. RNA from cells containing the U2dl-BD gene protects a fragment ~78 nt long, indicating that U2-bd RNA is 10–15 nucleotides shorter than wild type at the 3' end. This phenomenon is not due to a difference between the resident wild-type U2 genes and the cloned U2 gene we are using as a paradigm. Strains containing only a copy of the cloned wild-type gene have wild-type 3' ends (Fig. 3, lane *i*). We have also recovered the plasmid from yeast by transformation into *E. coli*, and confirmed that the sequence through the 3' end of the U2dl-BD gene is intact (not shown). Primer extension using the



**Fig. 4** Measurement of actin and *MATa1* pre-mRNA and mRNA levels in isogenic strains differing only at the U2 locus. Left panel, actin pre-mRNA and mRNA. Total RNA from each strain was annealed at 55 °C to a synthetic oligonucleotide (5'-ACACATAC-CAGAACCCTTATC-3') complementary to the second exon of the actin mRNA<sup>16,17</sup>. Top arrow indicates migration of complementary DNA derived from unspliced pre-mRNA. Bottom arrow indicates migration of cDNA derived from mature mRNA. Lane *a*, *rna11* strain grown at 23 °C; lane *b*, *rna11* strain shifted to 37 °C for 45 minutes; *c*, MA5c; haploid *MATa*, *his4-619*, *ura3-52*, *leu2-3,112*; *d*, MA5c1B: same as MA5c, except that it carries U2::*URA3* disruption in the chromosome and the YCpU2dl-BD centromere plasmid; *e*, 503d; *f*, HI2 (see Fig. 3 legend); *m*, markers, *HpaII* fragments of pUC13. Right panel, *MATa1* mRNA and pre-mRNA. Total RNA from each strain was annealed at 57 °C to a synthetic oligonucleotide (5'-TGATGTTGCTCTACTTTAGTC-3') complementary to the second exon of the *MATa1* mRNA<sup>18</sup>. Top two arrows indicate cDNAs derived from transcripts containing the first intron, bottom two arrows indicate cDNAs derived from transcripts from which the first intron has been removed. Lane *a*, *rna11* strain grown at 23 °C; *b*, *rna11* strain grown at 37 °C; *c*, MA2c; *MATa*, *his4-619*, *ura3-52*, *leu2-3,112*; *d*, MA2c1B: same as MA2c except contains the U2::*URA3* disruption in the chromosome and the YCpU2dl-BD centromere plasmid; *m*, markers, *HpaII* fragments of pUC13.

**Methods.** RNA from *rna11* strain ts382 (Yeast Genetic Stock Center) was used as a control to aid in identification of cDNAs derived from unspliced precursors. This mutant is a member of a class of mutants defective in splicing<sup>25</sup>. The expected cDNAs derived from unspliced pre-mRNA were either elevated relative to wild-type *RNA11* strains at permissive temperature, as in the case of actin, or accumulated upon shift to restrictive temperature, as in the case of *MATa1*. Isogenic strain pairs differing at the U2 locus were constructed by introducing YCpU2dl-BD into the strain, and then replacing the chromosomal U2 locus with the U2::*URA3* allele<sup>6</sup> by transformation with a linear DNA fragment<sup>26</sup>. As one end of the linear U2::*URA3* DNA fragment used for the transformation lies within the U2dl-BD deletion, this DNA is efficiently incorporated into the chromosome rather than the plasmid. To measure spliced and unspliced transcripts, total RNA (10 µg) was annealed to 0.2 ng oligonucleotide end-labelled with <sup>32</sup>P, in 16 µl 125 mM Tris-Cl, pH 8.3, 17.5 mM KCl for 30 min at the indicated temperature, and then cooled rapidly on ice. The reaction mixture was made up to 1.25 mM each dNTP, 10 mM MgCl<sub>2</sub>, 5 mM dithiothreitol, 40 µg ml<sup>-1</sup> actinomycin D, and 3–4 units of AMV reverse transcriptase (Life Sciences) in a volume of 20 µl and incubated for 30 min at 42 °C.

LI5 oligonucleotide revealed no differences at the 5' end of any of the U2 transcripts (not shown). We conclude that the shortened 3' end of the U2-bd transcript is a consequence of the internal deletion, and that the missing 3' end sequences are probably also dispensable to U2 function. Our data suggest a function for the dispensable sequences in snRNP assembly or RNA stability, because wild-type U2 seems to be preferentially accumulated in cells containing both wild-type and deletion U2 alleles (Fig. 3, lanes *d* and *g*), although a direct effect on 3' end formation is not excluded.

As the cells carrying only the U2dl-BD allele displayed no growth defects, we tested the effect of the U2 deletion on the splicing of both the yeast actin intron<sup>16,17</sup> and the first intron of the *MATa1* gene<sup>18</sup>. There was no detectable increase in the level of pre-mRNA and no significant decrease in the level of spliced mRNA in either case (Fig. 4).

The 210-nt yeast U2 may be folded to resemble other U2 RNAs (Fig. 1b). The highly conserved 5' 110 nt have the potential to form a pseudoknot<sup>19</sup>, like all other U2 RNAs<sup>14</sup>. The 3' end sequences can be folded to resemble stem III or stem IV (Fig. 1b, inset). We prefer the stem IV folding because trypanosome U2 (ref. 20) also seems to lack stem III (ref. 14). In this sense, the dispensable RNA in yeast U2 may be equivalent to the variable regions of ribosomal RNA<sup>8,9</sup> or RNaseP RNA<sup>10</sup>.

We thank Alan Weiner for support, Kivie Moldave for use of facilities, and Danielle Konings, Beth Shuster and Christine

Guthrie for communicating their unpublished results. This work was supported by institutional start-up funds (to M.A.) from the University of California at Santa Cruz.

Received 19 May; accepted 15 June 1988.

- Green, M. A. *Rev. Genet.* **20**, 671–708 (1986).
- Sharp, P. *Science* **235**, 766–771 (1987).
- Maniatis, T. & Reed, R. *Nature* **325**, 673–678 (1987).
- Parker, R., Siliciano, P. & Guthrie, C. *Cell* **49**, 229–239 (1987).
- Reddy, R. *Nucleic Acids Res.* **14**, suppl. r61–r72 (1986).
- Ares, M. *Cell* **47**, 49–59 (1986).
- Kretzner, L., Rymond, B. & Rosbash, R. *Cell* **50**, 593–602 (1987).
- Woese, C., Gutell, R., Gupta, R. & Noller, H. *Microbiol. Rev.* **47**, 621–669 (1983).
- Clark, C., Tague, B., Ware, V. & Gerbi, S. *Nucleic Acids Res.* **12**, 6197–6220 (1984).
- James, B., Olsen, G., Liu, J. & Pace, N. *Cell* **52**, 19–26 (1988).
- Siliciano, P., Jones, M. & Guthrie, C. *Science* **237**, 1484–1487 (1987).
- Patterson, B. & Guthrie, C. *Cell* **49**, 613–624 (1987).
- Siliciano, P., Brow, D., Roiha, H. & Guthrie, C. *Cell* **50**, 585–592 (1987).
- Ares, M. & Igel, A. in *Molecular Biology of RNA* (UCLA Symposium on Molecular and Cellular Biology Vol. 94 (ed. Cech, T.)) (Liss, New York, in the press).
- Fitzgerald-Hayes, M., Clarke, L. & Carbon, J. *Cell* **29**, 235–244 (1982).
- Gallwitz, D. & Sures, I. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2546–2550 (1980).
- Ng, R. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3912–3916 (1980).
- Miller, A. *EMBO J.* **3**, 1061–1065 (1984).
- Pleij, C., Rietveld, K. & Bosch, L. *Nucleic Acids Res.* **13**, 1717–1731 (1985).
- Tschudi, C., Richards, F. & Ullu, E. *Nucleic Acids Res.* **14**, 8893–8903 (1986).
- Maniatis, T., Fritsch, E. & Sambrook, J. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Riedel, N., Wise, J., Swerdlow, H., Mak, A. & Guthrie, C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8097–8101 (1986).
- Sanger, F., Nicklen, S. & Coulson, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5466 (1977).
- Boeke, J., LaCrute, F. & Fink, G. *Molec. Gen. Genet.* **197**, 345–346 (1984).
- Lustig, A., Lin, R.-J. & Abelson, J. *Cell* **47**, 953–963 (1986).
- Rothstein, R. *Meth. Enzym.* **101**, 202–211 (1983).



# Science product summer releases

*Some of the hottest science products introduced this summer are outlined this week, including a chromatophoresis instrument, an alternative to fetal bovine serum, and assays for plant hormones.*

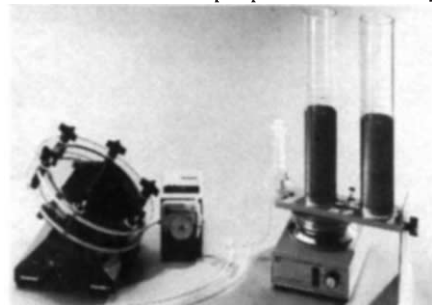
TECHNICAL Products International, Inc. has a **tissue sectioning system** that works as well with fresh tissue as it does with fixed tissue (*Reader Service No. 100*). The \$7,500 (US) Vibratome 2000 features a specimen bath that may be refrigerated



*The Vibratome 2000 tissue sectioning system.*

and circulated by a closed-loop built-in cooling system and pump. Technical Products International says uniform fresh or fixed tissue sections can be cut to any reasonable thickness — not restricted to 10  $\mu\text{m}$  increments — and the specimen height can be controlled with the push of a button. The Vibratome 2000 can be programmed to raise the stage by a pre-selected increment and cut a single section, or automatically to repeat the procedure for continuous serial sectioning.

Eppendorf has developed a relatively inexpensive approach to **cell fractionation** which relies on the properties of velocity



*Eppendorf's idea for cell fractionation.*

sedimentation and chamber reorientation (*Reader Service No. 101*). The Celsep 5440 consists of a peristaltic pump, a two-cylinder gradient maker and a 1,000-ml cell separation chamber. To use the Celsep, a reusable cushion solution and the gradient solution are loaded through the lower opening of the cell chamber. The cell suspension and an overlay solution are then added through the upper opening, and the separation chamber is reoriented horizontally so that the cells separate by sedimentation. After one to three hours, the chamber is once again inclined, and the cell populations can be fractionated in approximately 20 minutes. Eppendorf says the Celsep can hold 25-100 ml of a cell suspension which contains

1-20  $\times 10^5$  cells, and can separate cells which differ by more than 0.5  $\mu\text{m}$  in diameter.

Bionovus has simplified the **two-dimensional analysis of proteins** by coupling microbore high-performance liquid chromatography (HPLC) with gel electrophoresis to create a new technique that it has named chromatophoresis (*Reader Service No. 102*). The CP2000 Chromatophoresis System from Bionovus has a robotic transfer head that drives negatively charged proteins in the HPLC elution stream directly into a gel. The result is real-time analysis without loss of resolution, in much less time than it would take to perform both techniques separately, or to perform two-dimensional gel electrophoresis. Bionovus says the CP2000 can process up to 10 samples



*Chromatophoresis: new from Bionovus.*

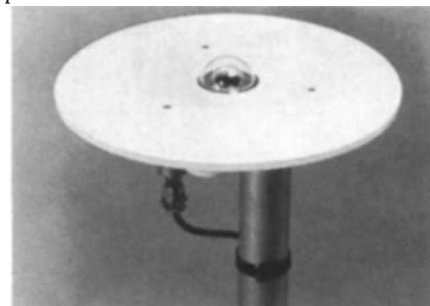
simultaneously, using different separation parameters, in less than 15 hours. An AT-class computer controls both dimensions of separation, the transfer process and the robotics. The CP2000 comes with a cassette and reagent kit, which includes a disposable microbore HPLC column, 10 prepackaged disposable gels, reagents, buffers and a computer diskette.

## Environmental extras

Joyce-Loebl Ltd has recently introduced its newest equipment for **remote sensing** studies — the Scandig series (*Reader Service No. 103*). Each set of instruments in

the Scandig series includes a single-pass scanner and film writing equipment. The Scandig range of colour drum scanning densitometers digitizes pictorial images with a resolution of down to 12.5  $\mu\text{m}$ , says Joyce-Loebl. The system is capable of scanning maps, photographs, line drawings and transparencies for conversion to digital form and subsequent input into image processors and computers. The product range includes both standard and large format devices, with scanning areas of between 300  $\times$  400 mm and 1,000  $\times$  1,200 mm. The Scandig colour film writers reproduce colour images from digital information onto photographic film.

Didcot Instrument Company Ltd provides two instruments for the **monitoring of solar radiation** — a solarimeter and an albedometer — both of which are based on type CM5 Kipp and Zonen cells (*Reader Service No. 104*). The £580 (UK) solarimeter operates using a Moll thermopile with a blackened receiver to measure



*One of Didcot's solar radiation meters.*

direct solar radiation. The receiver is protected by two optically ground and polished glass domes: an outer dome to protect against environmental factors, and an inner dome to reduce convective and radiative heat exchange between the receiver and the outer dome. Didcot says



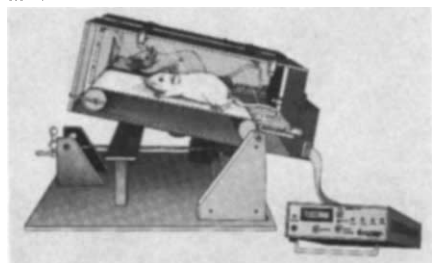
*The Scandig series of remote sensing equipment from Joyce-Loebl.*

its solarimeter measures accurately over the transmission band 0.3–2.5  $\mu\text{m}$ , with a response time of 70 per cent of the final reading within 3 sec, and an accuracy of  $\pm 1$  per cent. The £1,235 (UK) Didcot albedometer is based on two identical solarimeter cells, and can be used to record shortwave radiation balance and hours of sunshine, as well as total incident solar radiation.

### Animal studies

The Israel Institute for Biological Research offers an **animal behaviour laboratory system** for the assessment of the behavioural effects of drugs, chemical compounds and pollutants (*Reader Service No. 105*). The Supervisor system is designed to operate, control and collect data from up to eight behavioural boxes, based on a preset library of behavioural tasks designed to measure learning and memory, and anxiety and stress. The system collects two versions of data: a short, cumulative record for the daily evaluation of the experiment, and a long detailed record for in-depth analysis. The system records all relevant information for each animal, such as daily sessions and treatments. The system consists of a software package and an interface card, and is operable with levers, feeders, dippers, shockers and other components.

The latest animal research tool from Columbus Instruments is an **air-tight animal treadmill** for the collection of



For exercising or metabolic studies.

oxygen consumption data during exercise (*Reader Service No. 106*). The treadmill has isolated running lanes with separate electrical shock stimuli which can be configured for up to four rats or mice. The total volume of the chamber is approximately 15.4 litres, and it is compatible with any of the metabolic computers from Columbus Instruments. The basic £9,737 (UK) system includes a belt speed controller and an adjustable voltage shocker. The £13,487 (UK) fully automatic system can be programmed to run consecutive experiments with rest periods in between, and automatically monitors the number of trips to the shocker grid, time spent running, and time spent on the shocker grid.

### DNA details

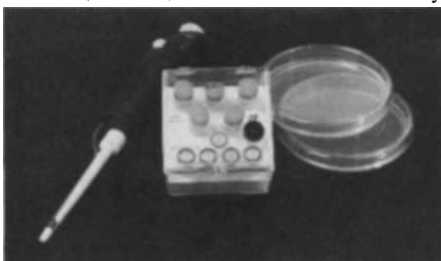
Micron Separations, Inc. has developed a **pre-mixed hybridization solution** for blotting, colony hybridizations and plaque



Hybridization solution in a pouch.

lifts called Hyb-IT (*Reader Service No. 107*). The company says Hyb-IT replaces the use of as many as ten solutions and reagents for the hybridization step, reducing preparation time and assuring more consistent test results. Prepared in a filtered 1 $\times$  concentration, Hyb-IT contains Denhardt's solution, SSC, 10 per cent dextran sulfate, herring sperm DNA, bovine serum albumin and buffers. It is available with 0 per cent or 50 per cent formamide for high and low temperature hybridizations, and intermediate concentrations may be prepared by mixing the two. Hyb-IT comes in a 50 ml resealable bottle for \$56–68 (US).

Invitrogen has announced the availability of a line of complete, fully optimized ***Escherichia coli* transformation kits** which offer the convenience of pre-aliquoted transformation media (*Reader Service No. 108*). The Ultracomp kits include enough fresh SOC media, competent *E. coli* and pre-aliquoted test plasmids for ten transformations. The kits are available with nine different strains of *E. coli*. DH1 $\alpha$ F', NM522, CJ236, JM109 and JM101 carry



Transformation kits from Invitrogen.

information for pili construction, allowing M13 infection. A combination of NM522 and CJ236 is used for site-directed mutagenesis, DH1 and DH1 $\alpha$  are used for routine cloning, and MC1061/P3 is used for transformation with vectors which require complementation with ampicillin and tetracycline genes carrying amber mutations. The Ultracomp kits come as

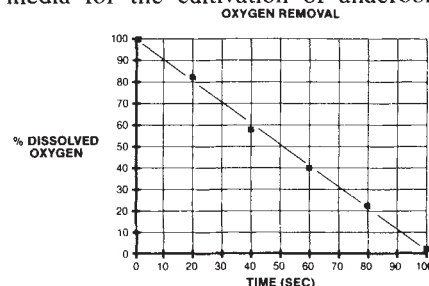
five aliquots of either 0.2 ml or 1 ml, and cost \$85–115 (US).

The German company SisWare offers **sequence analysis software** for the analysis of DNA, RNA and protein sequences (*Reader Service No. 109*). The DNA Analyser software package is written in Turbo Pascal language, and runs on an IBM PC or compatible. The program allows users to manipulate up to 32,000 nucleotides at the same time, and includes facilities for creating matrix plots with variable coordinates for nucleotide sequence comparisons, and for constructing hydrophobicity curves for protein sequences. With appropriate communications linkups, DNA Analyser performs sequence homology searches as well as searches for author names or keywords in both the EMBL and NRPD databases. SisWare says DNA Analyser can translate DNA sequences into all three reading frames in a few seconds, and can locate the cleavage sites of over one hundred restriction enzymes.

### Kits and reagents

Irvine Scientific has a standardized **substitute for fetal bovine serum** which it sells for between \$200 and \$250 (US) per litre — up to \$200 less than standard fetal bovine serum (*Reader Service No. 110*). The substitute is colostrum-free bovine serum, collected from calves slaughtered less than 24 hours after birth who have not been allowed to suckle. Irvine Scientific claims the chemistry of the serum is virtually identical to fetal serum, and is more standardized because it comes from animals of exactly the same age. The company tests lots of the serum for endotoxin and immunoglobulins using protein electrophoresis, and fills the serum containers under Class 100 sterile conditions.

The Oxyrase enzyme system from Oxyrase, Inc. is a biocatalyst for the removal of dissolved oxygen for the creation of **anaerobic environments** (*Reader Service No. 111*). Oxyrase is specific for dissolved oxygen, and does not exhibit the side reactions of other non-specific chemical reducing agents, according to the company. The company says as little as 1–2 ml of Oxyrase can prepare media for the cultivation of anaerobic



The rate of oxygen removal by one unit of the Oxyrase enzyme system under standard conditions.



microorganisms at a cost of as low as \$3 per litre. Oxyrase is available at 30 units per millilitre in 30 ml and 100 ml quantities for \$108 and \$300 (US), respectively.

Kits for the measurement of **plant hormones** are available from Idetek, Inc. (Reader Service No. 112). Idetek's Phyto-detek kits are enzyme immunoassays



Idetek's assays for plant hormones.

which contain monoclonal antibodies specific for indole-3-acetic acid, abscisic acid, *trans*-zeatinriboside, dihydrozeatinriboside and isopentenyladenosine. The company says the kits can detect the hormones at levels between 0.02 and 50 pmol per 0.1 ml in semi-purified extracts. Each \$300 (US) 96-well microplate kit includes tracer, tracer diluent, substrate, substrate diluent, wash solution, stopping reagent, plate sealers and a plate holder.

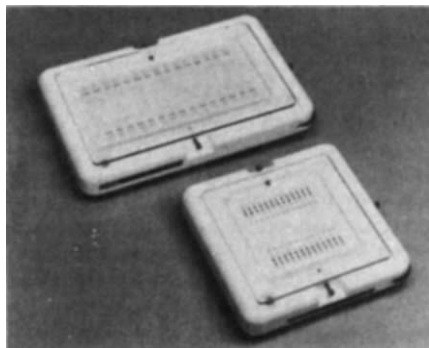
The fifth edition of *Linscott's Directory of Immunological and Biological Reagents* for 1988 and 1989 is now available (Reader Service No. 113). The **directory** lists sources for more than 26,000 different products in biomedical research, including more than 3,300 monoclonal antibodies, 9,000 conventional antisera and conjugates, and more than 14,000 other reagents such as lectins, recombinant DNA reagents, blood products, purified serum fractions, immunoassay kits, tissues, microbial cultures and antigens, separation media, enzymes, hormones, peptides, cell lines and laboratory animals. The author of the directory, William D. Linscott, has sold more than 18,000 copies of the first four editions of the directory, and plans to issue three supplements to update the current directory to the end of 1989. The directory costs \$55 (US) and £60 (UK).

Bachem now offer an ultra-pure **synthetic peptide** corresponding to the second conserved domain of gp120, the envelope protein of the human immunodeficiency virus (HIV) (Reader Service No. 114). The company says the peptide, designated C21E, was found to elicit antibodies which efficiently neutralized three different strains of HIV *in vitro*, but antibodies to the peptide did not inhibit binding of the virus to CD4+ cells. Bachem suggests that the region may be of importance in a post-binding event during virus penetration, and may be of use in the search for an

AIDS vaccine. The synthetic peptide costs \$125 (US) for 0.5 mg, and comes as a lyophilized powder.

### Thin-layer chromatography

Camag has a **horizontal developing chamber** that it says can be used to develop a high performance thin-layer chromatography plate from both sides towards the centre with any solvent, doubling the number of samples which can be analysed (Reader Service No. 115). The Camag chamber is made of PTFE and glass throughout, and can be operated in both sandwich and tank configuration,



Develops TLC plates from both sides.

with or without presaturation of the layer with solvent vapour.

### Lab housekeeping

Lambda Photometrics Ltd has developed a **laser safety shutter system** to allow the safe operation of lasers without the need for locked doors and special access keys (Reader Service No. 116). The safety shutter unit is positioned at the laser output window. In its "open" state, full laser power at beam diameters of up to 19 mm pass through without attenuation. When any access door to the area is opened, the shutter immediately closes, cutting off the beam completely, without interrupting or affecting the laser's operation, says the company. The shutter falls by gravity, not electricity, so the system protects safety even during a power failure. The system is composed of an electrically illuminated laser warning sign, a heavy-duty door switch and a master control unit.

The **Cyclon water distillation system** from Fistream has a patented condenser vapour trap which the company says eliminates all vapour-borne droplets, to ensure there is no carryover of impurities in the distillate (Reader Service No. 117). The Cyclon still features microprocessor controls, membrane touch switches, thermistor-controller safety systems, and an automatic boiler discharge to combat scale buildup in hard water. The Cyclon is offered with outputs of 4 litres and 8 litres per hour for a single distillation, and with a 4 litre per hour output for a double distillation.

### Setting it straight

The 14 April review of the model FT500 Fourier transform spectrometer from Chelsea Instruments was mistakenly printed along with a photograph of the company's model CI4000 protein sequencer (*Nature* 332, 666; 1988). The photograph was also misplaced by one paragraph, so that it appeared within a review of spectrometry software from Applied Photophysics.

A review of the Beckman Instruments model JE-5.0 elutriator rotor in the 2 June issue (*Nature* 333, 481; 1988) stated in error that the rotor separates living cells ranging in size from 2 µm to 5 µm. The rotor in fact separates cells of between 2 µm and 50 µm in size.

The £30,000 (UK) price quoted for the E. Leitz Cy-Scan cytology training system in the 26 May issue (*Nature* 333, 380; 1988) should have been given as £11,500 (UK). The higher price is for the Cy-Scan Laboratory Management System. □

*These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.*

### ADVERTISEMENTS

**Economical Modular Tissue Culture Incubator**

- Safe for AIDS research • Reliable, for In-Vitro fertilization
- Versatile enough for thousands of lab uses

**Easy to use** - Just flush with gas mixture, seal and place at approp. temp. Perform multiple experiments simultaneously w/out cross contamination. Thousands in use worldwide.

For information contact: **Flow Laboratories, Ltd.** or, **Billups-Rothenberg, Inc.** • P.O. Box 977  
Del Mar, CA 92014-0977 USA • (619) 755-3309

### Transfer of DNA/RNA to solid supports, Southern/Northern blots and filter hybridization



Our accurately temperature-controlled rotisserie oven gives **important advantages** as against bag: 2.4-3 ml hybridization solution. Lower background. Higher intensity of bands. Multiple hybridizations possible. Uniform distribution of solution.

**Bachofer GmbH P.O. Box 7089  
D-7410 Reutlingen W-Germany**